

A comprehensive test-ban now, please

The case for a comprehensive test-ban treaty, necessary for global security and civility, is made more pointed by worries about proliferation. But now the British seem to be dragging their feet.

WHAT should be the next step in the process of strategic arms control that has powerfully, if occasionally, lifted spirits in the past two decades? The important article by Lord Zuckerman on page 392 shows what has been achieved, often against the prevailing odds, to reduce the risks of global nuclear war. The partial test-ban treaty of 1963, in whose negotiation the British played a leading part — remember Prime Minister Harold Macmillan's return from Moscow in a white fur hat? — and in which Zuckerman's personal role was central, would have been necessary on environmental grounds. Zuckerman now argues that it is time to complete those distant negotiations to arrive at a comprehensive test-ban treaty (CTB).

None of this belittles the importance of what has been achieved since 1963. A string of treaties now restricts the capacity of great powers to wage nuclear war. The most remarkable is the Anti-Ballistic Missile Treaty, predicated on the counter-intuitive view that the world would be safer if the nuclear powers were not defended against attacks by strategic missiles. At least on paper, there is now also a prospect of substantial reductions in the numbers of strategic missiles deployed (see *Nature* 361, 1; 1993). Yet all that, in changed circumstances, does not suffice for stability. A CTB is the logical next step.

For more than three decades, since technical talks on the feasibility of a CTB began on the coat-tails of the second United Nations Conference on the Peaceful Uses of Atomic Energy in 1958, the CTB has been waiting in the wings. The first serious negotiation, marked by an Anglo-American and Soviet moratorium on testing lasting more than two years, was ostensibly defeated by uncertainties about the remote detection of underground explosions by seismic means; the determination of the military to improve the efficiency of their warheads was still more influential. Then, towards the end of Jimmy Carter's US presidency, there were high hopes that a CTB would materialize. But that treaty, which would have been signed by Britain, the Soviet Union and the United States, would not have met the present need, while Carter's successor, President Ronald Reagan, would not have fought for its ratification in the US Senate.

Circumstances are now different. Zuckerman remarks on the consequences of the fragmentation of the Soviet Union into several independent republics: those with nuclear weapons on their soil are potentially nuclear powers. (The Ukraine, by some accounts, is negotiating with the West to trade this potential for more urgent needs, cash among them.) Since

1958, France and China have emerged as nuclear powers while, more recently, countries as different (or, rather, as similar) as India, Pakistan, Iraq and North Korea have acquired the reputation of nursing nuclear ambitions. To make sense now, a CTB must be genuinely comprehensive.

That objective has been made more urgent by the impending review conference of the Non-Proliferation Treaty (NPT) in 1995, when the parties to the treaty will have to decide whether it continues. Previous review conferences have been rowdy occasions, partly because the nuclear powers have not been able to report much progress in strategic arms control. Two years from now, there will be more to boast about, but the nuclear powers themselves will be alarmed at the difficulties of policing the NPT. The safeguards regime operated by the International Atomic Energy Agency is more effective than its extreme critics say, but it is cumbersome. If signatories of the NPT were also signatories of the CTB, policing would be a lot simpler. So the established nuclear powers now have an interest in a CTB that transcends the eagerness of their military to keep on testing weapons.

Except, it seems, in Britain. Zuckerman, as loyal a public servant as the British government could find, is plainly dismayed that the British have declined to follow France, Russia and the United States in the current moratorium on nuclear testing. They are, of course, in an odd position, having nurtured plans to deploy Trident missiles in submarines and an air-launched nuclear weapon still in the early stages of development. They say they need to test the warheads for these weapons, but President Bill Clinton says the Nevada test-site to which the British have access will be shut in 1996. Will it not be ironical if the government that rightly boasts of having been the driving force behind the Partial Test-Ban Treaty is seen to drag its feet on the CTB for the sake of strategic plans irrelevant to its present military needs? The British government should read Zuckerman and then change tack. □

Europe's research

A meeting of research council heads will put pressure on Brussels and may give Strasbourg a bigger role.

THE arrival in Brussels of Antonio Ruberti as the new member of the European Commission responsible for research has been widely welcomed. He is adroit, energetic



and knows first-hand what science is about. He says (see *Nature* **361**, 286; 1993) that he hopes to mould the evolution of the European Communities (EC) research programme by taking fuller account of the needs of working scientists. That makes sense. Too much of what the EC has previously attempted, often at the whim of individual commissioners, has naively been more concerned with industrial than research policy. Ruberti will be more, not less, effective now that some of those projects have been transferred elsewhere in the commission.

But how is the EC to get closer to research? And what part does the EC have to play in the development of research in Europe as a whole? It will not have escaped attention in Brussels that, on 18 and 19 January, the administrative (not the political) heads of Europe's research councils held an informal meeting at Bonn to brood about these questions among others. Although the only tangible decision was that there should be another meeting later in the year, more generally the sense of the meeting seems to have been that the pan-European research enterprise is too important to be left exclusively to the EC.

The research council heads are powerful fellows, who collectively spend on research nearly a hundred times as much as the EC. Already, through their support of the European Science Foundation (ESF) in Strasbourg, they have a European research agency of their own which, significantly, embraces parts of Europe where the EC's writ does not yet run. (The next secretary-general of ESF, from July, will be Dr Peter Fricker, for many years the director of the Swiss national research foundation.) In ESF and from their collective experience, they also have a vivid appreciation of how, in science, the aspirations of individual researchers help to determine the pattern of research — just the link Ruberti says he wants to establish in the EC's research programme.

The research councils can plainly help Ruberti a great deal, not least with advice. He is, after all, the first research commissioner to include education in his brief; one of his tasks should be to use a substantial part of the EC's research money to stiffen research support for higher education, especially in parts of the community where science is comparatively neglected. But the research council heads, at Bonn, had by all accounts set their vision on a wider Europe. In particular, they were talking of the possibility that ESF might become "a common instrument" for the improvement of European science. That is an exciting prospect, which Ruberti could do much to encourage. □

Watching watchdogs

The Chemical Weapons Convention will work only if verification procedures are sensitive to public scrutiny.

THE Chemical Weapons Convention, signed in Paris last month by more than 130 countries after 25 years of negotiation in the United Nations, is due to come into full operation within two to three years, but already there is alarm over

arrangements for policing it. In Britain, the responsibility will fall to the Department of Trade and Industry (DTI), one of whose tasks is to promote the welfare of British industry. The treaty will require that governments should restrict the export of equipment that has both military and civil applications. But only last year, DTI was one of several government departments caught up in the Matrix-Churchill affair, when an ex-government minister admitted that officials had been encouraged to be less than scrupulous in applying that same 'dual use' test to exports for Iraq. The record suggests that the nominated custodians of British responsibilities under the treaty are not above suspicion.

Overall responsibility for ensuring compliance with the convention will rest with a new body, the Organisation for the Prohibition of Chemical Weapons, to be established at The Hague. A preparatory commission is meeting in the Netherlands next week to decide how that, and in particular its intended technical secretariat, will function. But the convention also imputes a significant role to national authorities, which will be responsible for reporting on national compliance, for example by collecting information on the activities of the chemical industry.

In Britain, the national authority is being set up by DTI. In one sense, the choice is logical; much of the information to be collected will be of a nature that the industry would ordinarily prefer to keep to itself. That is particularly so of production and marketing data on 'dual use' chemicals which, although widely used for civilian purposes, can also be the precursors of chemical weapons. Who better, industrial officials argue, to protect it against foreign predators than DTI?

But, as with any weapons control treaty, the effectiveness of the convention will hang on signatories' confidence and trust in the verification procedures. There is a clear conflict between the industry's need to protect confidential information and the public interest that such information is subjected to independent critical scrutiny. Not to put too fine a point upon it, can DTI alone be trusted to do the job required?

The solution is — or should be — relatively straightforward. The national authority should consist not only of government and industry officials but also of independent members, perhaps drawn from the scientific and academic communities. Short of full public transparency — agreed to be impractical — the presence of individuals able to raise awkward questions and voice suspicions about potentially illicit activities may be the best guarantee that the integrity of the review process is observed, and that self-interest is not allowed to overcome due caution.

Such procedures have already worked well for genetic engineering. But so far the president of the Board of Trade, Michael Heseltine, the minister responsible, has not been receptive. Answering a parliamentary question last week, he merely stated that the composition of the national authority was still under discussion. But with the credibility of an international treaty at stake, this is one area in which John Major's commitment to open government should mean that the financial interests of the few are not allowed excessive influence over the security interests of the many. □

Cambridge and Heidelberg compete for new European gene database

London & Munich. Two of the world's oldest university towns are bidding to host Europe's main store of information about human and animal genes. The databank, to be known as the European Bioinformatics Institute (EBI), is likely to become a cornerstone of the biotechnology industry and will compete with the more advanced US GenBank.

Germany wants the EBI to go into a new 2,400 square metre building on the European Molecular Biology Laboratory (EMBL) site in Heidelberg. Details are being kept secret until next week's deadline for tender, although rumours suggest that the state of Baden-Württemberg would donate the land and the federal government would provide DM12 million (US\$7.5 million) in capital.

But British scientists have put together an £8 million (US\$12 million) proposal for the institute to be located close to Cambridge. Backed by the Medical Research Council (MRC) and the Wellcome Trust, which would each contribute half of the initial capital costs, the institute would be established on a 55-acre parkland site nine miles south of Cambridge.

At the moment, EMBL operates Europe's gene and protein databases (the European Nucleotide Sequence Database and SWISSPROT, which is run jointly with the University of Geneva) and several other specialist databases on a budget of DM5 million a year and a staff of 40. The Commission of the European Communities (EC) has also donated a total of ECU2 million (US\$2.6 million) in the past four years.

The 13-year-old EMBL data library has outgrown its base, and EMBL has decided to move it to a separate institute that would include a research department. The department would keep the library's technology up to date and provide support and training for users.

EMBL estimates that EBI will cost more than DM12 million. Although a call for tenders was put out in December for the building and equipping of the institute as an EMBL outstation, a decision about running costs was deferred until next month's meeting of the EMBL council. Europe lags behind the US GenBank, which offers free service internationally, a more sophisticated computer system and an active research team. In addition, Europe wants its own database to protect itself if the US government ever begins charging for use of GenBank.

Member countries had only two months to submit a bid, a deadline favouring those already well along. Although Spain and

Sweden have expressed verbal interest, only Germany and Britain have developed specific proposals.

The site will be chosen on the basis of its access to expertise in molecular biology and computing and its links to fast electronic networks. Ideally, for administrative reasons, it should also be associated with a laboratory with the same international status as EMBL.

British officials point out that the Cambridge site offers direct access to the many molecular biologists working at both the university and the MRC's Laboratory of Molecular Biology (LMB), as well as the Agricultural Research Council's Institute of Animal Physiology and Genetics Research at Babraham. Those working on techniques for interpreting information about gene and protein structure from data stored in the database would have similar access to experts in computing.

"Cambridge has an enormous research community of exactly the type that we want the EBI to flourish in", says John Sulston of the LMB, who has recently been appointed director of the Sanger Centre, a new institution for genomic research. Cambridge also has excellent high-speed data links with the rest of the world. A 2-megabit link available through the JANET network is soon to be upgraded to a 34-megabit link through the new network SUPERJANET.

The estate on which the EBI and the Sanger Centre would be based, Hinxton Park, is being bought by the Wellcome Trust. At present it includes not only an eighteenth-century house, but also 4,000 square metres of laboratory space for scientists at the Sanger Centre to use while waiting for their own building to be completed. The initial capital costs, which would be shared by Wellcome and the MRC, include £2 million to transfer the existing databank from Heidelberg.

At a meeting in London last week, the proposal to locate the EBI in Cambridge received the support of both the Science and Engineering Research Council and the Agri-

cultural and Food Research Council.

Germany's alternative bid to locate the EBI at Heidelberg offers its own advantages. The current data library would move only across the road to what is now a parking lot, and EMBL's status as an international laboratory would simplify administrative problems. The institute would also benefit



Britain has proposed Hinxton Hall (above), outside Cambridge, as the site for the European Bioinformatics Institute; Germany has Heidelberg (right) in mind.



from the talent already assembled in Heidelberg.

Counting against it, however, are Germany's poor electronic datalinks with the rest of the world. Germany has always lagged behind other high-technology nations in electronic networks and, according to Graham Cameron, EBI project leader, this weakness has caused quite a few problems in the past couple of years. The immediate future offers the possibility of upgrading its current 64K bit links to a 2 megabit commercial link — at a price. But anything faster lies far into the future.

The choice of a site and a decision on how to meet operating costs will be made at next month's EMBL council meeting. One option is to reject the concept of EBI as an EMBL outstation and instead set it up as an independent foundation.

David Dickson & Alison Abbott

Massey will leave NSF amid controversy over its future

Washington. The decision last week by Walter Massey, director of the US National Science Foundation (NSF), to become second-in-command at the University of California (UC) system leaves the Clinton administration with the difficult job of finding someone who can reconcile competing demands on NSF from the politicians who fund it and from the scientists who receive the money.

Massey, who has accepted the position of provost and senior vice president for academic affairs at the \$7-billion-a-year, 166,000-student UC system, joined NSF in March 1991 from the University of Chicago. His appointment is expected to be ratified on 18 February by the UC Board of Regents, and he is expected to leave NSF by the end of March, less than halfway through his six-year term as NSF director.

Massey, 54, is a physicist turned administrator with a reputation for working behind the scenes. Yet his tenure at NSF, which this year will spend \$2 billion on academic research, was marked by a growing public debate on what role NSF should play in strengthening the US economy. That debate has been sharpened by one of the few promises about research that President Bill Clinton made during his campaign, namely to balance the civilian and military portions of the government's \$75-billion budget for research and development, which now tilts 58:42 towards defence.

Massey waded into that debate last summer by forming a commission to study the future of NSF. His hand was forced by demands from Congress that NSF support more research relevant to industry. Although the commission (see *Nature* **360**, 285; 1992) issued a ringing endorsement of basic research and reinforced previous comments from industrial leaders that NSF should continue to train scientists and not try to pick which technologies have the greatest commercial potential, university researchers were greatly upset that NSF was even thinking about abandoning its original mission.

Such an idea infuriates NSF officials, who insist that Massey never intended to turn away from investigator-initiated basic research. But they are quick to say that NSF must change with the times or risk losing its central role in US science and, more importantly, any hope of a larger budget.

"Two years ago we weren't even thinking about where we were going", says Mary Anne Fox, a professor of chemistry at the University of Texas at Austin and a member of the NSF commission as well as of the National Science Board, the 24-member governing body for NSF. "He recognized

that change was coming and he began to pull the NSF back into line."

But many would disagree, saying instead that Massey has accelerated a shift away from individual researchers and towards large-scale and commercially orientated activities that began under his predecessor, Erich Bloch. What NSF needs, they say, is an eminent scientist with the skills to persuade the public and the politicians of the benefits of basic research. Anything less, they argue, will lead to a further deterioration of an already bad situation in many fields in which hundreds of talented young scientists compete for a single university position and fewer than one in four grant applications are funded.

Massey's departure coincides with the reauthorization this year of legislation that broadly spells out the types of programmes that NSF should be supporting. This exercise, conducted every five years, involves three congressional committees, one in the House of Representatives and two in the Senate (because of overlapping jurisdiction). Their interests, although different, will also influence the continuing debate about NSF's direction.

The topics to be addressed include the health of the research enterprise, in particular whether the system is too large for the funds available; the need for billions of



Massey is heading west.

dollars to renovate ageing research facilities on campus; the sorry state of science and mathematics education in elementary and secondary schools; NSF's role in promoting a dozen or more 'critical technologies' and its ability to work more closely with industry on government-funded generic research. The next director will necessarily spend a great deal of time reconciling divergent opinions on these and other issues.

Nevertheless, the job is an attractive one. Because NSF is an independent agency, the director has an opportunity to take his case directly to senior White House staff, including the budget director, and on rare occasion even to the president himself. Vice President Al Gore and the new science adviser, John Gibbons, are expected to play major roles in the selection process. **Jeffrey Mervis**

Italians criticize selection process

Munich. The new president of Italy's research council (CNR) takes up his position this month amid controversy over the selection process. Microbiologist Enrico Garaci,



Enrico Garaci

currently rector of the Second University of Rome, benefited from the discredited 'Lottizzazione' system, in which positions are divided among the main coalition parties in government. In Garaci's case, the post fell to the Christian Democrats because the minister of research, Sandro Fontana, belongs to that party.

Garaci's appointment comes as public pressure grows to end the Lottizzazione system. Physicist Carlo Bernardini from the University of Rome was the first to speak out publicly against a system that he describes as

"a clearly illegal procedure which happens in no other democracy". Garaci himself was a victim of Lottizzazione when he stood for election in 1989: although he received the most votes, the Socialists were given the powerful position of Mayor of Rome. As a result, Garaci was owed a favour, which has now been granted.

Despite unhappiness about the system, few criticize his qualifications for the job. He is familiar with the organization as a scientist and member of one of the CNR's 15 advisory committees (biology and medicine), and his ten years as rector of the Second University of Rome have given him plenty of administrative experience.

Garaci supports the prevailing opinion that Italian science must move away from a system in which state funding is allocated evenly (everyone gets a small grant) and toward a focus on the best research. He believes that the grant reviewing system, currently handled by several advisory committees, should be standardized and that referees from other countries should be used.

Alison Abbott

Gibbons sees Gore links as asset for science

Washington. The new director of the White House Office of Science and Technology Policy (OSTP) hopes that Vice President Al Gore's interest in the subject will give science a greater voice in shaping the policies of the Clinton administration. And he suggested that President Bill Clinton may take a different approach on such megaprojects as the Superconducting Super Collider (SSC) and Space Station Freedom.

John Gibbons was last week sworn in as director of OSTP after the Senate confirmed him unanimously (his appointment on Christmas Eve as the president's science adviser took effect immediately). Earlier in the week, during his confirmation hearing before the Senate Committee on Commerce, Science and Transportation, Gibbons gave hints of the new president's thinking on a number of important issues facing the scientific community.

The creation of a civilian agency modelled on the successful Defense Advance Research Projects Agency in the Pentagon, at one point a major plank in Clinton's campaign platform to strengthen US high-technology industries, no longer appears to be under consideration. Instead, Gibbons said in response to questioning, "the most opportune way is to fully utilize what we do now". That means yet another review of the multi-billion-dollar national laboratories, many created to fight the Cold War and now looking for civilian missions, increased support for joint projects with industry and a much larger budget for the National Institutes of Standards and Technology, home to the fledgling Technology Administration.

Yet even such an innocuous approach as greater collaboration faces potential hazards. Are there any restrictions that the administration should place on companies to prevent them from taking their knowledge overseas and using foreign labour to make the advanced products, wondered Senator John Kerry (Democrat, Massachusetts). Gibbons parried the question by saying that the Clinton administration would look at the reasons behind such US corporate flight, as well as why many young high-technology US companies must look for help overseas rather than within the United States. "I think that the jobs will stay here if our people are productive and well-trained", he said.

But Clinton is also eager to reexamine current policies in important areas, Gibbons asserted. In particular, Gibbons suggested that the \$8.5-billion SSC and \$30-billion space station would receive much closer scrutiny as the president looks for ways to reduce the \$300-billion federal deficit. The failure to obtain significant contributions to the SSC from other countries (see *Nature* 361, 288; 1993) can no longer be ignored,

he said. Similarly, he added, the space station includes certain technologies "that should be international from the outset". Both projects are expected to come under attack this year from Congress, and their future will be precarious if the new president is less supportive of them than George Bush.

D. Allan Bromley, Gibbons' predecessor, put great stock in a series of interagency committees that reviewed US efforts in sev-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

John Gibbons, left, talks with Sen. John Warner before his confirmation hearing last week.

eral scientific fields. Gibbons believes that such panels, called Federal Coordinating Committees for Science, Engineering and Technology (FCCSET), "have been effective in some instances" but that more can be done with them. He said he hoped to benefit from a close relationship with Gore because "the science adviser's ability to convene a group [such as FCCSET] is much less than the vice president's". He also said that he

expects to put greater emphasis on environmental issues and that one of his four deputies will be given that portfolio among other duties.

Although it has been more than a dozen years since Gibbons worked at a university (he directed a centre on energy and the environment at the University of Tennessee before becoming director of the congressional Office of Technology Assessment in 1979), he sprang to their defence in response to questions about federal investigations of conflict of interest, misuse of research dollars and scientific misconduct. Gibbons said that universities should be treated differently from industry because "we have to remember that they aren't out there making money — in fact, they often lose it.... I hope that we remember they are our partners, not our adversaries, in achieving national goals".

What a science adviser thinks matters little, of course, unless he has the ear of his bosses in the White House. Thanks to his status as a member of the newly created National Economic Council, Gibbons spent last weekend with Clinton and his cabinet at the presidential retreat of Camp David, 90 miles from Washington, mapping out an economic strategy. Although his presence is seen as a positive sign that Clinton intends to incorporate science into his deliberations, a more significant measure will come in mid-March when the president submits his first budget.

Jeffrey Mervis

Grace period urged for patents

London. Pressure is growing on the British government to give research scientists a one-year 'grace period' after the publication of research findings in which to apply for a patent on their results. The change from the present policy of having to apply before publication would bring UK policies in line with those in the United States.

The latest support for this proposal, which has already been recommended by the Science and Engineering Research Council, comes from the government's top science advisory committee, which also suggests that all biomedical research students be given training in the importance and evaluation of intellectual property rights. Both recommendations are made in a report* published this week.

The recommendations are intended to help address what a committee chaired by Peter Doyle, director of research and technology for ICI plc, describes as a "development gap" between research laboratories and industrial exploitation — a gap which, the

committee says, is a "particular difficulty" in Britain. One suggested remedy is the creation of a new programme, MEDILINK, by the Department of Trade and Industry, to increase "towards market" research and development through projects linking academic and industrial researchers.

The current system for managing intellectual property in British universities is "underdeveloped and under-resourced", according to scientists contacted by the committee, and pressure to publish research quickly "has restricted the interaction between academia and industry." To avoid this — and presumably to give industry a chance to identify patentable discoveries in the scientific literature — it suggests legislation allowing both the British and the European Patent Offices to grant a one-year grace period.

David Dickson

* A Report on Medical Research and Health. Advisory Council on Science and Technology. Office of Science and Technology. HMSO. £13.00

IBM research to stress products and results

Washington. As International Business Machines (IBM) Corporation fights to reverse one of the worst corporate declines in financial history, its \$6.6-billion research programme is shrinking and changing direction. A cut of \$1 billion in the overall programme, most of which goes into product development, is being accompanied by a shift towards product-related work and reduced support for risky outside ventures.

IBM, the world's largest computer manufacturer, had an abysmal year in 1992. The share price of its stock dropped by half, it lost \$5 billion and, for the first time ever, it laid off an estimated 25,000 people in a work force of 300,000. James McGroddy, head of IBM research, says his division will be hiring fewer scientists and engineers and that the work force of 3,300 people at IBM's three research centres in New York, California and Switzerland will shrink this year, from attrition, by "a couple of per cent".

The company's research division, which has sponsored Nobel prize-winning basic research in atomic physics, materials science and other fields along with its more product-related research, spent about \$650 million last year. Although that figure will drop by as much as ten per cent this year, McGroddy says that most scientists will still have what they need. "We're coming from a background where we've lived pretty well", he says.

What will change is the way in which the money is spent. In recent years, IBM has shifted from basic science and technology towards work that is more customer-orientated, what McGroddy calls "services, applications and solutions". Such work accounts for about five per cent of the research division's current activities, but McGroddy expects it to reach 20 per cent by the end of 1994.

Although IBM is reducing its support for semiconductor technology and related sciences, the change "doesn't mean there is no opportunity for physicists", says McGroddy. But they may wind up doing something other than physics, he says.

The changes at IBM are the latest example of an ailing high-technology company trying to tie its research programme more directly to the bottom line. Frank Nayadas of the Sloan Foundation, a former IBM vice president, says, "What IBM will have to do is focus its research in the businesses it expects to be in", even if that means a shift away from hard sciences and towards such

areas as software and algorithm development. But McGroddy welcomes that shift. "The research lab is not a national monument you don't ever want to change", he says.

IBM's problems are also likely to mean the end of support for unprofitable outside ventures. This new attitude has already claimed one victim, supercomputer designer Steve Chen, whose Supercomputing Systems Inc. (SSI) of Eau Claire, Wisconsin, closed its doors on 25 January after IBM, its

dwindling for vector processors of the kind Chen was building. "I'm not sure his machine is what the world needs", says Curt Rohrman, a Wall Street analyst with First Boston Corporation. In fact, IBM was this week expected to introduce a different kind of supercomputer, a massively parallel processor based on years of in-house research.

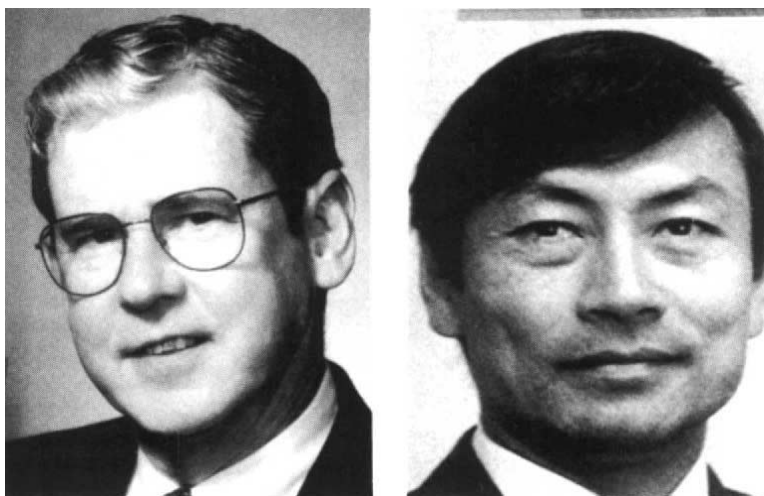
IBM may also become more stingy with its funding of university research. The research division spends \$4 million a year supporting university research in areas of interest to the company, and other divisions have similar arrangements. Ettore Infante, a vice president for academic affairs for the University of Minnesota system, which now gets \$700,000 a year from IBM, expects some programmes to suffer as a result of the retrenchment.

University researchers may also be affected if, as expected, IBM becomes more aggressive in seeking funding from the US government. The Navy, for example, has funded IBM laser scientists, and the company belongs to a consortium

of companies working on superconductors with money from the Pentagon's Defense Advanced Research Projects Agency. Federal grants make up only a tiny part of the company's total research budget, says IBM spokesman Denis Arvay, but any increase would be felt by academic scientists competing for the same funds.

IBM insists that its commitment to basic research will remain strong, but acknowledges that its future is uncertain. The motto of the research division is to be both "famous for its science and technology" and "vital to IBM", says McGroddy, but it appears that the latter has now taken on greater importance.

Tony Reichhardt



McGroddy's research division will shrink; IBM cuts ties to Chen, right.

sole backer, completed a five-year investment in the small company, which was attempting to build the world's fastest supercomputer. IBM is said to have spent more than \$100 million on SSI before it decided that the company would never make the grade.

"[Chen's] machine was a very beautiful design", says Malvin Kalos, director of the Cornell Theory Centre. "It could have been extraordinary if it had come to market very early." But Chen had worked on the new supercomputer for five years with nothing to show for it, and was believed to need another two years.

Even if he had succeeded, the market is

Cray's Rollwagen gets Commerce job

Washington. John Rollwagen, chairman of Cray Research, Inc., has been nominated by US President Bill Clinton for the number two job at the Department of Commerce under Ron Brown. As head of the world's largest seller of supercomputers, Rollwagen adds the technological and business experience that Brown lacks to a department that Clinton wants to figure prominently in his programme of economic stimulation.

Rollwagen, CEO of Cray since 1980, is expected to be a popular choice among high-technology industrialists. Charles

Larson of the Industrial Research Institute says he "certainly understands the technology and the problems of commercializing technology". Rollwagen is currently a member of the presidential Advisory Committee for Trade Negotiations.

Rollwagen talks regularly with Vice President Al Gore about high-technology issues, and last summer Gore made a campaign visit to Cray's headquarters outside Minneapolis. Gore is well known for his interest in creating a national "information infrastructure".

Tony Reichhardt

Britain plans broad strategy on genome, approves therapy

London. Keen to exploit Britain's strengths in molecular biology and genetics, the government is planning a national strategy for human genome research, drawing on the skills and interests of university scientists, research councils, medical charities and the pharmaceutical industry. Last week's announcement came as approval was granted for the country's first gene-therapy experiments.

As a first step towards a national programme, the Office of Science and Technology has commissioned a broad-ranging report on the state of genome research in



Kay Davies

Britain, to be delivered by the end of the year. The report will cover not only research and training related to the human genome, but also the scope of (and presumably barriers to) commercialization of genome-related technologies and the ethical questions involved.

The report will be written by a small group of experts chaired by Kay Davies, recently appointed director of the Medical Research Council (MRC)'s new Clinical Sciences Centre in London. Other members of the review group will be announced after the first meeting on 5 February of an Advisory Committee on Human Genome Research, which has just been set up by William Waldegrave, the cabinet minister responsible for science.

Sir William Stewart, the government's chief scientific adviser, will head the advisory committee. In addition to the heads of two research councils and two charitable foundations (the Wellcome Trust and the Imperial Cancer Research Fund), the advisory committee will include the research

directors of four of Britain's largest pharmaceutical companies: Glaxo, ICI, SmithKline Beecham and the Wellcome Foundation.

In announcing the creation of the advisory committee, Waldegrave emphasized the way in which genome research provides what he described as "the basis for the most modern techniques of gene therapy". Coincidentally, a research team headed by Roland Levinsky, head of the University of London's Institute of Child Health, received on the same day the go-ahead for the experimental transplantation of a functioning adenosine deaminase (ADA) gene into the bone marrow cells of a child suffering from ADA deficiency.

Similar experiments have already been carried out in both the United States and Italy but in those cases, only T cells were altered and the process had to be repeated at regular intervals. In contrast, the British technique will involve the manipulation of stem cells as well as T cells which, if successful, will result in one of the first permanent implantations of a manipulated human gene.

The ADA experiments are proceeding after the decision by a government advisory committee on gene therapy, chaired by Sir Cecil Clothier, that it had no ethical objections to the experiment, although it emphasized the need for close monitoring of the patient's progress.

Also last week, the MRC announced that, given the increasing interest in gene therapy — one of the principal aims of the new Clinical Sciences Centre is to stimulate work in this field — it is setting up a steering committee to ensure that such work is carried out responsibly. In particular, the steering committee will be responsible for ensuring that clinical trials of gene therapy conducted by MRC scientists are "well planned and properly conducted" as well as encouraging interaction with industry.

Those keen to proceed with gene-therapy experiments, however, are still waiting official endorsement from the government. This had been expected last week (see *Nature* 361, 285; 1993) in the form of a decision to set up a permanent body to oversee this work, as proposed by the Clothier committee last year, but has apparently been delayed for what a government spokesman described as "administrative reasons".

The government's response is still thought to be imminent. "It is time that we had a response", Dai Rees, secretary to the MRC, said last week. "Things are moving ahead fast in this field, and there is a danger that the government could start to hold things up."

David Dickson

INDIA IN BRIEF

New Delhi. The Indian government has stopped all field releases of an insect introduced to combat a noxious weed because it has begun to feed on a major crop. Eight years ago, the government imported the Mexican beetle *Zygogramma bicolorata* to control the weed *Parthenium hysterophorus*, which accompanied grain imported from the United States in the 1950s, but last year the insect was found feeding on sunflower, from which oilseed is derived. Tests on the beetle's affinity to 200 indigenous plants, done before the experiment was launched, are being repeated, but there is also the possibility that a species closely related to *Z. bicolorata* and known to be a sunflower pest elsewhere in the world was unwittingly imported in the original batch from Mexico.

■ India plans to proceed with large-scale trials of a birth-control vaccine that the World Health Organization (WHO) says is unsafe. The Indian National Institute of Immunology, which developed the vaccine, says that the criticism is rooted in a desire by WHO to promote its own vaccine, developed in the West. Both vaccines aim to neutralize human chorionic gonadotropin (hCG), produced at the time of conception and essential for development of the fetus. The Indian vaccine uses purified beta subunit of hCG as an antigen, which WHO officials say can be harmful to healthy women because it also reacts to luteinizing hormone (LH). The WHO vaccine is a synthetic peptide representing a small fragment of the beta hCG that does not react with LH. A phase-3 trial on 10,000 women is planned using slow-release capsules intended to give protection against pregnancy for as long as two years.

■ India has decided to drop its attempt to buy a Cray YMP supercomputer after the United States imposed what India calls "unacceptable" conditions on the sale. The deal, proposed in 1988, soured largely because US officials impose greater limitations on exports to countries that have not signed the Nuclear Non-Proliferation Treaty; in India's case, those included having a person from Cray onsite at all times and restricted physical access to the machine itself. The Indian Institute of Science in Bangalore, which intended to use the supercomputer for basic research in crystallography and biology, has already spent \$1 million on new facilities to handle the machine. C. R. Rao, director of the institute, calls the rejection "a blessing in disguise" and says the institute will instead install its own massively parallel processing computer, called PARAM, with workstations throughout the campus.

K.S. Jayaraman

British VAT blocks loan of Japanese computers

Tokyo & London. A bizarre application of Britain's value added tax (VAT) may scuttle a loan from Japan to Imperial College in London of two component computers from the fifth-generation computer system developed by the Institute for New Generation Computer Technology (ICOT) in Tokyo.

ICOT has offered to provide two ¥10-million (US\$80,000) Personal Sequential Inference (PSIs) machines free. ICOT would like to encourage worldwide development and application of its fifth-generation computer, in which Japan has invested more than ¥50,000 million over the past 11 years. But under VAT rules, either the college or ICOT would have to pay VAT of ¥3.4 million (£17,500) and customs duties of ¥900,000 on the machines.

This is not the first time VAT has hampered attempts by Japan to provide support for British research (see right). In 1991, the institute offered to lend David Warren of the University of Bristol the two PSIs and the necessary software. ICOT and Warren planned to pay the VAT bill out of a research grant from the Department of Trade and Industry (DTI). However, the DTI wanted the university to get half of the amount from British companies and would not allow ICOT's contribution to count towards this. When the university was unable to meet the DTI's requirements, "to our great disappointment" the deal fell through, says Kazuo Narita, managing researcher of ICOT.

Last June, ICOT received another proposal from John Darlington and Keith Clarke

of Imperial College, London, both of whom had been involved in assessing the fifth-generation project. They have suggested overcoming the VAT problem by making the loan through the Imperial College Computing Forum, which would then lend the computers to the university. As a VAT-registered organization requiring the tax to be paid on subscriptions received from private companies which are members, the forum is in a position to reclaim any VAT paid on the Japanese computers.

However, the refund could be delayed, arriving too late to be used for research. Although Clarke says that this temporary shortage could be recouped from the corporate subscriptions paid into the computing forum, Narita thinks that any arrangement for the current financial year "unfortunately has to be given up".

ICOT has provided PSI machines to the National Institutes of Health (NIH), Argonne National Laboratory and the University of Oregon in the United States and to the Australian National University. Some of the machines, for example those at NIH, have not been put to much use because they are difficult for non-experts to handle (see *Nature* 356, 273; 1992), but ICOT has had "absolutely no problems" with tax or import duties in the United States or Australia, Narita says.

The Japanese institute plans to make a third attempt at cooperation with Britain in the next fiscal year (starting on 1 April).

David Swinbanks & David Dickson



UK muon source given exemption

Tokyo & London. ICOT is not the first Japanese organization to face a VAT bill for providing support to British research. The Institute of Physical and Chemical Research (RIKEN) near Tokyo thought that it might have to pay about £2 million (US\$3.2 million) after agreeing to invest £15 million in Britain's Rutherford Appleton Laboratory (RAL) to build and run a pulsed muon source. RIKEN was able to avoid most of the VAT only by calling the facility a 'gift' from the Japanese government.

In 1990, RIKEN and RAL, with the blessing of the British government, agreed to construct a muon source that would tap RAL's powerful pulsed proton beam. Construction and operating costs would be paid by Japan's Science and Technology Agency (see *Nature* 354, 98; 1991).

Late in 1990, Japanese and British officials realized that they faced a bill of about ¥400 million from the British Treasury and customs for VAT (17 per cent) and customs duty (7–10 per cent) on the ¥1,600 million (£8.3 million) cost of construction of the facility, which RIKEN planned to build in Japan and transport to Britain. Such a tax bill would have caused "severe damage" to the joint project, says Katsutoshi Yasufuku, director for international cooperation at RIKEN.

Plans were altered so that some of the heavy equipment, such as shielding, costing about ¥600 million, would be built in Britain. That would avoid some import duties, but not VAT. Finally, in November, most of the construction costs were exempted, as was the equipment being supplied by Japan, leaving a tax of ¥50 million on British construction materials.

The special concession on import duty was granted on the grounds that the equipment was a gift from the Japanese government and because the project is for nonprofit, scientific purposes. But it seems unlikely that this procedure will be applied to the case of ICOT (see left above).

David Swinbanks & David Dickson

Meetings on British science

NATURE is organizing two one-day public meetings to discuss proposals for the forthcoming white paper on the organization of British science, in Edinburgh and London.

On **Friday 19 February**, there will be a meeting in **Edinburgh**, jointly sponsored by the Edinburgh International Science Festival. The speakers will be Professor David Glover (University of Dundee), Sir Graham Hills, Professor J Maxwell Irvine (University of Aberdeen) and Professor David Wallace (University of Edinburgh). The meeting will be at the Queen Mother Conference Centre, Royal College of Physicians, 9 Queen Street, Edinburgh, and will start at 9.30 a.m.

Admission will be by ticket, free of charge, obtainable from Bruce Durie, Edinburgh International Science Festival, 1 Broughton Market, Edinburgh EN3 6NU. Coffee and tea will be provided. There will also be a sandwich lunch for those who want it, at a cost of £5: please send a cheque, made out to the Edinburgh International Science Festival, with your ticket application.

On **Friday 19 March**, there will be a meeting at 9:30 a.m. in **London** at the Royal Society, 6 Carlton House Terrace, SW1. The speakers will include Sir Eric Ash (Rector, Imperial College London), Professor Michael Brady (University of Oxford) and Sir Mark Richmond (Chairman, Science and Engineering Research Council).

Admission will be by ticket, free of charge, obtainable from Mary Sheehan, *Nature*, 4 Little Essex Street, London WC2R 3LF. Coffee and tea will be provided. There will also be a sandwich lunch for those who want it, at a cost of £5: please send a cheque, made out to *Nature*, with your ticket application.

US panel disagrees on cancer testing method

Washington. A US National Research Council committee on risk assessment methodology is divided over the continued use of a controversial testing procedure in which animals are exposed to massive doses of a chemical to determine whether it causes cancer.

In a report* released last week, the committee acknowledged limitations in the information provided by bioassays that use the maximum tolerated dose (MTD), the highest dose of a chemical compound that can be administered to a test animal without seriously affecting its health from causes other than cancer. But a majority of its members agreed that more data are needed before MTD testing can be replaced with a radically new approach.

Six of the committee's 17 members disagreed with that conclusion and called for a new approach to carcinogenicity testing. The committee minority recommended that the MTD bioassay be replaced as the initial test of a chemical's carcinogenicity with studies aimed at determining the physiological and metabolic mechanisms by which a chemical causes toxicity in animals. A panel of experts would then set dose levels for the animal bioassays based on these data — an approach that, the minority argues, would yield results more relevant to humans.

The use of the MTD has long been controversial. Critics argue that the sensitivity level for MTD testing has been set too high and that the number of chemicals tested by this method and found to be carcinogenic is so large that attention is being focused unnecessarily on chemicals that pose only insignificant risk to humans.

"The screen is supposed to narrow down the materials on which you would focus further attention", says Richard Reitz of the Dow Chemical Company, a committee member who sides with the minority. Reitz says that it is not very useful when about half the materials tested by this method are identified as animal carcinogens.

It is also argued that MTD bioassays provide little information on what happens at low doses and that at the MTD some chemicals induce cancer through mechanisms that do not occur at lower doses. "When we have mechanistic data at the molecular level, these data should be the driving force behind the long-term carcinogenicity studies", says Michael Gallo, associate dean for research at the Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, and a committee member who also holds the minority view. Gallo says that he hopes the report is not read as an endorsement of "business as usual for another decade".

Opposing Gallo and Reitz is Bernard Goldstein, director of the Environmental and Occupational Health Sciences Institute

at the Robert Wood Johnson Medical School and Rutgers University, and chair of the committee. MTD is needed because of a lack of a basic understanding on how chemicals work and the absence of appropriate testing procedures, he says.

Despite this disagreement, the majority recommended that the MTD should continue to be one of the doses used in carcinogenicity bioassays, although other doses downward from MTD/2 possibly to MTD/10 or less should be included. The report states "that there are as yet no validated mechanisms of carcinogenicity that support lowering the MTD and that failing to use the MTD for carcinogen screening will reduce the sensitivity of the bioassay and diminish the opportunity to compare results among chemicals and species". The majority believes that no additional tests

related to carcinogenicity would normally be required if bioassays that include the MTD do not produce cancers in test animals.

Developed in the late 1960s and early 1970s, animal-testing methods using the MTD have been standard in the United States for more than 15 years. In these bioassays, groups of 50 or more animals are usually tested at two or more doses of a chemical, with the highest dose being the MTD. Because the animals are exposed to the chemical for most of their lives, rodents with relatively short lifespans of two to three years are generally the preferred test species.

Diane Gershon

* *Issues in Risk Assessment Committee on Risk Assessment Methodology*, Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Academy Press, Washington DC, 1993 (ISBN 0-309-04786-2).

William Petty papers purchased for £1 million

London. The collected papers of one of the most influential but least familiar participants in the scientific revolution of the seventeenth century, the economist and polymath William Petty, have been purchased by the British Library for more than £1 million — a record sum for a historical figure of the period.

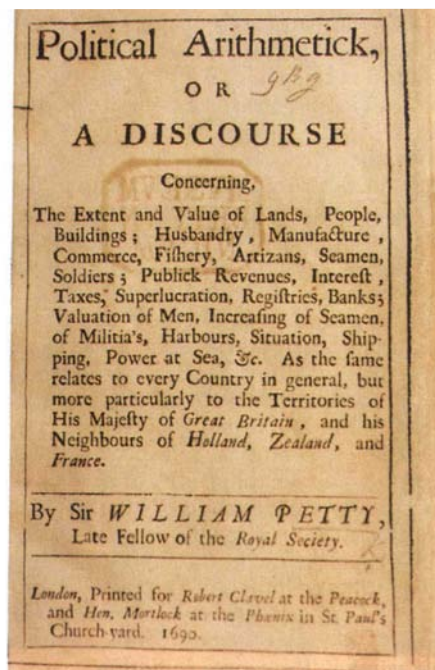
Petty is credited as the inventor of the science of statistics and is best known for his studies of the population of seventeenth-century London. Such work has led individuals ranging from Karl Marx to Maynard Keynes to regard him as the father of political economy.

But Petty, who trained as a doctor and later became a founding member of Britain's Royal Society, was also an intensely practical man. His ground-breaking survey of Ireland for Oliver Cromwell, resulting in 103 large-scale maps known as the "Down Survey", is seen as one of the milestones in modern cartography (as well as one of the first practical applications of the "scientific method" described by Francis Bacon). He also outlined plans for the first catamaran — which he described as a "double-bottom" boat.

Petty has remained in the shadows partly because his descendants have allowed scholars very restricted access to his papers. The British Library received a grant of £700,000 from the National Heritage Memorial Fund to buy the papers from his descendant, Lord Shelburne. The total cost is thought to be about £1 million.

Peter Barber, a map specialist at the library, says that the price paid for the Petty papers "reflects the importance that we attach to them". They are expected to throw light on both the origins of economic theory and statistics, as well as on the history of Anglo-Irish relations.

David Dickson



Simon Walker/The Times

Long-term national databases

SIR — Your article on North Sea data and “privatization” (*Nature* 361, 101; 1992) captures the general flavour of the debate although I would not wish to claim that all government departments of necessity take a short-term view, for many do not. However, in the case of the North Sea, the Department of Trade and Industry (DTI) appears to be taking (or is having to take) a short-term view of a long-term problem of curation of a major national geological database.

It is impossible to overstate the scientific and commercial value of the oil and gas database for the North Sea, which has cost billions of pounds to acquire and millions of pounds to curate. It will continue to be valuable long after oil exploration has ceased in the North Sea, just as geological data collected 150 or more years ago by the British Geological Survey (BGS) are still used. Other issues include the confidential access to the database, and the impartial advice that the survey can provide because of its status as a non-profit-making organization, but the crux of the issue is the quality and long-term continuity of the curation. Of course it will be claimed that contracts can be written to ensure that curation is undertaken to current BGS standards — just as new software commonly claims to be uniformly compatible! Of course companies will claim

that they can provide a service that is just as good as that provided by the BGS and perhaps they will claim that they could provide long-term continuity of curation. But the fact is that the BGS has existed as the national survey for more than 150 years and I would expect it to do so for another 150. How many geological companies have this track record or this future expectation? But this is the sort of timescale over which national datasets must be maintained and this is why I believe that the potential saving of a few pounds in 1993 is insufficient to justify jeopardizing the integrity of a dataset that will be consulted by geoscientists in 2093 or even 2193. Just as the United Kingdom environment must be sustained for the benefit of future generations, so must we also ensure the sustainability of the UK knowledge base for those following us.

The geological database for the North Sea, lodged with the DTI but obtained through the efforts of the oil and gas exploration industry over the past 25 years, can only be satisfactorily curated on a long-term basis by the BGS. In the national interest the data should remain with the British Geological Survey.

Peter J. Cook

*British Geological Survey,
Kingsley Dunham Centre,
Keyworth, Nottingham NG12 5GG, UK*

Diet and cancer

SIR — A recent leading article in *Nature* (359, 760; 1992) stated that further studies of the possible relation between fat consumption and breast cancer incidence are unwarranted in view of the negative results obtained in several case-control studies, notably the recent ones by Kushi and Willett.

We believe that this judgement is premature. Cross-cultural and migration studies have provided evidence for a relationship between dietary fat intake and breast cancer, and that this effect requires a generation to become evident. This time lapse is in line with the apparent long latency of breast cancer and points to a major potential flaw in case-control studies. Such studies have the inherent limitation that they reflect current dietary patterns and become less accurate when recollections of past dietary habits are required. None of the cited case-control studies provided dependable data on fat intake before the age of 30 and thus did not address the issue of the dietary environment at puberty and shortly thereafter, a putative window for cancer initiation.

The role of incorporated body fat

resulting from increased fat consumption is also ignored. It has been shown that increased body fat decreases oestradiol 2-hydroxylation resulting in decreased excretion of 2-hydroxyoestrone, but increases the formation of 16-hydroxylated metabolites; one of these, 16 α -hydroxyoestrone, has been shown to be genotoxic while no such effects were observed for 2-hydroxyoestrone. These changes can be detected in cross-over studies in which large differences in fat intake are induced. Unlike these studies involving major alterations in fat consumption, in conventional case-control studies dealing with relatively small differences in fat intake the relationship between fat intake and breast cancer risk may not therefore be apparent.

We therefore suggest that the issue of the effect of fat intake on breast cancer needs to be assessed by properly designed epidemiological and biological studies that take into consideration the above interpretative pitfalls.

H. Leon Bradlow

Jack Fishman

*Strang-Cornell Cancer
Research Laboratory,
510 East 73 Street,
New York, New York 10021, USA*

Columbus a Jew?

SIR — The “Copernican” side of Columbus’s enterprise (*Nature* 359, 564; 1992) seems to have been generally underestimated, if not ignored. But this link could open new perspectives on the origins of modern science.

Certainly, the “primitive skill in the measurement of longitude” displayed by Columbus cannot be simply accounted for as a purely practical ability. It rather suggests that Columbus may have known more about astronomy than the background of a simple mariner would have allowed him. Indeed, it seems likely that he benefited from the work of skilled Jewish astronomers and mathematicians, particularly his contemporaries Abraham Zacuto and José Vizinho. The members of the Jewish community, in their particular social condition, could pursue more unfettered scientific activity in such fields as astronomy and cartography, taking advantage of Arab discoveries and getting close to some of the subsequent Copernican views. Moreover, they were stimulated to make practical use of scientific knowledge, in response to the persecution to which they were subjected. The problem is then to establish in what circles some kind of new scientific ideas started to spread, with that obvious “prudence” that you rightly refer to.

A good case for the Jewish origin of Columbus himself has been made by Simon Wiesenthal in his very interesting book *Operation Neue Welt — Judenverfolgung und Columbus-Reise* (Laffont, Paris, 1972/1991). Curiously enough, this claim has been recently strengthened by the suggestion that Pope Innocent VIII (1484–92) may have had an important role in providing Columbus with more than financial support (Ruggero Marino, *Cristoforo Colombo e il Papa tradito*, Newton Compton, Rome, 1991). Innocent VIII was certainly from a Jewish family (his father was called Aharon Cybo), and he had close connections with Lorenzo de’ Medici (which included financial enterprises too). It is possible to trace links between Copernicus and people connected in some way with Florence and the Medici family and between the same circle and the discovery of the New World.

By analysing in depth the cultural Jewish background of Columbus, and the international relations of the Jewish community, we may come to understand what would otherwise appear as a strange case of mysterious scientific clairvoyance.

Umberto Bartocci

*Dipartimento di Matematica,
Università,
06100 Perugia,
Italy*

Harvesting of whales

SIR — Professor A. S. Blix is reported (*Nature* 360, 523; 1992) as saying that “any resource that can be harvested on a sustainable basis . . . should be harvested . . . Not doing so is a waste of resources.” Professor L. Walløe expressed a similar view in a public debate with me recently in Stockholm. If their view were to prevail, we could say goodbye to all national and regional parks and to the greater part of the planet’s wildlife and flora.

A more enlightened — not to say civilized — view, which is in accord with the decisions of the UN Conference on Environment and Development, is that if wild animals or plants are harvested, then such harvesting should be limited so that it is sustainable. Furthermore, according to both the UN Convention on the Law of the Sea and the Agenda 21 document from UNCED, “living marine resources” that are already depleted to below their maximum potential level of productivity must be permitted to increase to that level. Such restoration can be achieved by restraining any catches to be less than the currently sustainable amount, which can include making catches temporarily zero. Where there is great uncertainty about the catch rates that might be sustainable — as there is with whales — then wise management calls for more cautious quota decisions.

Those are the elements of the policy of the International Whaling Commission (IWC) and, increasingly, that of governments and regional international organizations seeking to restore overfished fish stocks to higher levels of sustainable productivity. Decisions for temporary zero catch quotas in such conditions have nothing whatsoever to do with whether “... [whale] populations are in a particularly precarious state . . .” or whether a particular catch number will or will not pose “... a serious threat to the remaining population.”

Then, there is virtually no scientific or political “... dispute over the algorithm . . .” for calculating catch quotas for minke whales in the northeast Atlantic. The present difficulty arises from the fact that the IWC Scientific Committee has agreed, unanimously, that these whales declined in number continuously from at least 1953 to 1983 (the last year for which there are appropriate data) under the impact of Norwegian whaling. This means that “restoration” is mandatory under international policies and rules. So quotas set by the IWC would be cautious and, specifically, less than Norwegian whalers want for economic reasons and have been promised by Walløe and other Norwegian officials.

That is the main reason the Norwegian government has unilaterally decided to allow its whalers to resume next year. Public statements that the national quotas have been calculated using the IWC algorithm should be taken with a large pinch of salt.

You report other (unnamed) Norwegian scientists as feeling that the “generous” allocations for lethal research on whales have more to do with politics than with science. In this connection it is perhaps worth noting that the Norwegian government has just decided to make a roughly equivalent sum available for propaganda to “explain” why Norway has decided to isolate itself from the rest of the world on all matters — scientific, managerial and political — concerning whaling.

Sidney J. Holt

*International Fund for Animal Welfare,
Podere Il Falco,
06062 Città della Pieve (PG), Italy*

SIR — For the past three years I have been studying the biology of minke whales off the west coast of Scotland where, off the Hebridean island of Mull, a small whale-watching tourism industry is developing. The whales in this area belong to what the International Whaling Commission calls the ‘North East Atlantic Stock’. The Norwegian government has announced plans to resume whaling next year on this same stock when the animals are feeding in the Barents Sea, off Spitzbergen, and along the west coast of Norway.

In this context, I should like to correct the figures reported in an “Alternative view on whaling” (*Nature* 360, 523; 1992). The estimate quoted of about 80,000 minke whales (actually 86,700) is for the northeast Atlantic, not the world. The 95 per cent confidence limits for this estimate range from 58,000 to 115,000. There has never been a “previous estimate” as low as 18,000. For several years, Norwegian scientists insisted on a figure of 113,000 obtained from marking experiments (I. Christensen & K. J. Rørvik, *Rep. Int. Whal. Comm.* 31, 255–257; 1981). Then in 1983 they agreed with the IWC Scientific Committee on 66,000, and later agreed to a revised figure of 60,000, with a similar confidence limit range. Corresponding pre-exploitation figures are calculated simply by adding catches to estimates of current numbers, with some correction for births and natural deaths. For this reason, estimates of depletion are more robust than estimates of stock number. Since 1976, the IWC has classed stocks less than 54 per cent of their pre-exploitation number as “Protection

Stocks” with automatic zero catch quotas. The IWC resolution of 1991 endorsing a revised procedure for calculating catch quotas explicitly reaffirmed that there must be no catching from stocks which would have been classed as “Protection Stocks” under the old rules. Such stocks include the northeast Atlantic stock which was so classified in 1985. There are no significant differences between the estimates of the mid-1980s and the recent ones. Improved calculations presented to the IWC Scientific Committee in 1992 strongly confirm that the catch limit should remain at zero.

Vassili Papastavrou

*University of Bristol,
Department of Zoology,
School of Biological Sciences,
Woodland Road, Bristol BS8 1UC, UK*

Culture deposits

SIR — Some years ago, the story goes, on receiving a refusal to a request for a certain virus, the enterprising petitioner succeeded in isolating the virus from the envelope that enclosed the reply. Most workers make newly described microbes accessible and deposit living cultures in a publicly available collection. This is important to avoid losing the microbe and to enable others to verify the published observations.

Some people do refuse to make specimens available, however, because of the fear of being overtaken by others. Such unpleasant people are few and should not be allowed to impede free science. Unfortunately, their cause is facilitated by one of the rules of the International Code of Botanical Nomenclature¹. This rule applies to yeasts and other fungi and stipulates that only preserved, not living, specimens of newly described species need be deposited. A preserved specimen of a new species is useless for most purposes. Furthermore, many of its characteristics, given in the published description, are unverifiable. Examples of such taxonomically valid, yet unverifiable, species include two yeasts, *Filobasidiella depauperata*² and *Myxozyma sirexii*³.

A rule enabling the publication of observations that cannot be confirmed is anti-scientific. It should be changed, so that the deposit of living specimens of all new species and their accessibility is a condition of accepting the validity of the species.

J. A. Barnett

*School of Biological Sciences,
University of East Anglia,
Norwich NR4 7TJ, UK*

1. *International Code of Botanical Nomenclature* (Koeltz, Königstein, 1988).

2. Samson, R.A., Stalpers, J.A. & Weijman, A.C.M. *Antonie van Leeuwenhoek* 49, 447–456 (1983).

3. Spaaij, F., Weber, G. & van der Walt, J.P. *System. appl. Microbiol.* 15, 427–431 (1992).

Prospects for a comprehensive test ban

Lord Zuckerman

Progress towards an international moratorium on the testing and development of nuclear weapons has been slow and tortuous. The time is now ripe for a treaty to be agreed.

DURING the final weeks of George Bush's presidency, we were constantly being informed of his determination to crown his reign with a treaty that would dramatically cut the number of intercontinental nuclear warheads held by the United States and the former Soviet Union. He succeeded. He and Boris Yeltsin called for the number to be reduced from about 25,000 to 3,000 a side. They knew that this could not be achieved overnight, and neither was dreaming of a nuclear-free world. A fraction of those they would still retain would be enough to devastate utterly their respective homelands.

The treaty document they have signed is, of course, only a beginning. On both sides, and particularly on Yeltsin's, a lengthy process of ratification has still to follow. Four of the confederation of independent republics into which the old Soviet Union has transformed itself possess nuclear weapons, and apart from Russia (which has inherited the Soviet Union's position), none of the others has as yet even formally acceded to the Non-Proliferation Treaty (NPT). However, what the Bush/Yeltsin agreement does signal is a recognition that the nuclear arms race has long since passed the point where it made any political, strategic or even military sense. That recognition takes us back to the very beginnings of the race in the 1950s and to the attempt to ban any testing of warheads.

The build-up

So far, the only effective move towards sanity concerning nuclear weapons was in 1987 when, in a remarkable meeting of minds, Mikhail Gorbachev led Ronald Reagan to agree that intermediate- and short-range ballistic missiles should cease to form part of US and Soviet nuclear arsenals. The resulting Intermediate Nuclear Forces (INF) Treaty and, more important, the issue of mutually balanced force reductions, had been under vigorous and bitter discussion for

years but, in practice, the reciprocal build-up of nuclear weaponry had come to be regarded as inevitable, given the hostility and enormous political divide between Western and Warsaw Pact powers.

Suspicion that one side or the other might launch a lightning strike had not only made a mockery of most, if indeed not all, previous efforts to decelerate the nuclear arms race, but also, paradoxically, had imbued it with additional momentum. For example, the US-Soviet ABM (anti-ballistic missile) Treaty of 1972 placed a strict limit on the development and deployment of ABM systems lest they destabilize the East/West balance of nuclear power. Simultaneously, however, it gave an enormous boost to the development of missiles that carried several independently targeted warheads (MIRVs) — and it did this despite the 'interim' agreement reached in the same year by the two sides (SALT I) — to limit the size of their strategic nuclear arsenals. Weapons designers had engineered themselves into a position in which they enjoyed what in effect was a free run in the presumed effort to counter each other in the interests of national security, regardless of rational political, economic and strategic considerations.

The ultimate fantasy came in 1983, when the US Congress agreed to endorse Reagan's Strategic Defense Initiative (SDI), to embark on the development of a space-based defence system — the mirage of a 'defensive astrodome' that no ballistically launched warhead could penetrate — and so, to quote Reagan, "render nuclear weapons impotent and obsolete". To its protagonists and to the industrial interests concerned, the elusive prize was worth every cent that would be called for. The arguments that exposed the strategic futility of the whole undertaking were left unanswered, and despite the sequence of technical setbacks that for some 10 years now has characterized the research and development programme, SDI goes on

being funded. In the 1993 US military budget of \$274.5 billion, no less than \$3.8 billion has been allocated to SDI.

The hard facts

Even so, the continuing SDI fantasy has not weakened the tide of nuclear disarmament now running. The atmosphere of trust that began in 1987 triggered an avalanche of change in military and world affairs that no one could have foreseen. How much was due to Gorbachev's meeting with Reagan in Reykjavik, how much to his parallel domestic policies of *glasnost* and *perestroika* and how much to economic considerations will be for future historians to decide.

All that we know now is that what had been an ostensibly firm Union of Soviet Socialist Republics has disintegrated into a loose and turbulent Commonwealth of Independent States, together with three separate sovereign Baltic states. So deep-seated are the social, ethnic, political and economic problems that have followed the collapse of the previous *dirigiste* communist government that it is a matter for speculation whether Yeltsin, who little more than a year ago upturned Gorbachev for not moving fast enough, will now be able to withstand the widespread reaction of those — including the military and security services — who are experiencing the hardships occasioned by a too-rapid economic and political revolution. Equally, one can only speculate whether the tide of disarmament would be slowed were Yeltsin to fall. Nor is it yet possible to guess how the parallel turmoil and bloodshed marking the dissolution of the Warsaw Pact organization will end, or to foresee what kind of Europe will emerge from the chaos that can partly be attributed to the economic difficulties attendant on the reunification of Germany.

Barely a year after his INF agreement with Reagan, Gorbachev not only gave notice of the end of central communist control of the Soviet Union but also announced that he had ordered major unilateral troop reductions. The cold war was to end. Since then, major mutual reductions of conventionally armed forces in Europe have been agreed. At the same time, the START negotiations that had been grinding on since the early 1980s, with the aim of agreeing reductions in intercontinental nuclear armaments, were abruptly concluded at the

INTERNATIONAL ARMS-CONTROL TREATIES

	Treaty	Signed	Ratified	Signatories
PTB	Partial Test-Ban Treaty	1963	1963	UK, US, USSR (now 119)
NPT	Non-Proliferation of Nuclear Weapons	1968	1970	146 states
ABM	Anti-Ballistic Missile Treaty	1972	1972	US, USSR
INF	Intermediate Nuclear Forces (Europe)	1987	1987	US, USSR
START I }	Reduction and Limitation	1991	—	US, USSR
START II }	of Strategic Offensive Arms	1993	—	US, Russia

end of July 1991, when Gorbachev and Bush agreed that the maximum number of delivery vehicles on either side should not exceed 1,600 — quickly reduced to 1,250 — with warheads to 6,000.

This agreement was ratified by the US Senate despite the fact that there is considerable uncertainty about the issue of the central control of all nuclear armaments in the old Soviet Union, in four of whose now-independent republics (Russia, Kazakhstan, Byelorussia and the Ukraine, of which only the first two have so far ratified the START I Treaty) nuclear weapons are deployed. Gorbachev was ready to agree to anything that needed to be done to reduce the nuclear menace, real or unreal, as perceived by the United States. Thus the Krasnyogorsk radar station, whose erection the United States had claimed was in breach of the ABM Treaty, would be demolished. The Soviet Union would withdraw its long-standing insistence that the British and French strategic nuclear weapons should be included in the Western total, as well as its claim that the SDI programme constituted an impediment within the framework of intercontinental nuclear parity. Both sides also agreed to start withdrawing to their own territory the so-called tactical (battlefield) nuclear weapons deployed beyond their own borders.

Despite the experience gained in implementing the 1987 INF Treaty, disposing of the tens of thousands of warheads covered by the START agreements will take years. Fortunately, Russia and the United States have the same interest in preventing proliferation, and cooperation in the process of dismantling their nuclear arsenals will need to be firmly organized if mutual trust is to be sustained through the 1990s. Prevention of sales of nuclear materials to would-be nuclear-weapon states, or to terrorist organizations, and surreptitious restoration of 'downloaded' warheads to the missiles from which they were removed, must be assured by mutual monitoring of stockpiles by the two sides.

Why nuclear tests?

A comprehensive ban on nuclear tests has now become an equally urgent issue, partly because it is bound to be raised when the NPT comes up for review in 1995, and partly because it raises the question of the safety of stockpiled warheads waiting to be disassembled. For both these reasons, in June 1990 the relevant armed services committees of the US Congress appointed a three-man panel of physicists under the chairmanship of Sidney Drell to make an independent survey of the safety of stockpiled weapons (Drell was until recently co-director of the Stanford Center for International Security and Arms

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Yesterday's men plan for tomorrow — Gorbachev and Reagan at Reykjavik.

Control: the other two members were John Foster Jr, an ex-director of the Lawrence Livermore nuclear establishment and of Defense Research and Engineering in the Pentagon, and Charles Townes, a Nobel laureate in physics).

The panel's specific terms of reference were 'to assess the risk of "accidental detonations" due to any cause, including aircraft crashes and fire, that could result in the dispersal of plutonium into the environment, or which might "even generate a nuclear yield"'. Its report, published in December 1990, recommended that older warheads should be made equal in safety to the newest designs which incorporate the 'enhanced' nuclear detonation safety system (ENDS) whose purpose is to prevent premature arming; the use of insensitive high-explosive instead of normal high-explosive for the outer warhead shell; and the need for the warhead's central core of plutonium to be in a fire-resistant pit to prevent its dispersal were the weapon to be involved in an aircraft fuel fire following on a crash. The panel suggested that in the interests of safety it might sometimes be necessary or useful to retest warheads that had been rebuilt.

The congressional committees concerned consequently arranged that R. E. Kidder of the Livermore Laboratory, a senior physicist who had assisted in previous congressional inquiries into nuclear matters, should undertake a 'peer review' of the Drell report. In particular, before the United States agreed to be party to a "highly restrictive or comprehensive test ban environment", they wanted an estimate of the number and yields of underground tests that might be needed to ensure the safety of altered warheads. Kidder's answer was 10 to 20, if all the warheads of three US ballistic weapons (Minuteman III and Tridents C4 and D5), which according to present plans will remain in service, were modified to accord with modern safety stan-

dards. He endorsed the Drell recommendations, emphasizing (as did the Drell panel) the hazardous proximity to the detonatable third-stage rocket propellant of the warheads of the two types of Trident missile, C4 and D5, now deployed. In September 1992 in *Arms Control Today*, Kidder even stated that there are "no significant technical reasons why the US nuclear stockpile cannot be rapidly prepared for a CTB [Comprehensive Test Ban], with little if any additional nuclear testing".

In subsequent testimony before a panel of the House of Representatives Armed Services Committee (31 March 1992), Drell gave as his firm view that the additional and marginal gain in safety that could be derived from continued testing was less important than "the political value, real or perceived, of a CTBT for strengthening or even preserving a non-proliferation regime". Here he was echoing the unequivocal opinion of Richard Garwin, a physicist who has been closely concerned with nuclear weaponry for more than 40 years and who, in testimony to the Senate's Committee on Foreign Affairs on 27 February 1992, declared that he had no doubt "that stockpile reliability, safety and security can be maintained without nuclear testing". He warned at the same time that the practical problems involved in stockpiling warheads to await dismantling and for the safe storage of plutonium, depleted uranium and tritium, are considerable. The production of new nuclear materials should cease immediately for, as Garwin put it, we live in an age of "nuclear surplus". The US government soon came to the same conclusion. In September 1992 a 'new production reactor', which was to have had the sole function of producing tritium for nuclear stockpile, was cancelled.

The views about testing expressed by Drell, Kidder and Garwin also seem to have been accepted by the Congress's

Armed Services Committee. In October 1991, the then Soviet Union declared a one-year moratorium on nuclear tests — now extended by Yeltsin to mid-1993. Within days, the House of Representatives' Democratic Leader tabled a Nuclear Testing Moratorium Act, which after a few weeks was introduced to the Senate. A year later, in early September 1992, the Senate approved a motion to suspend tests for 9 months, and to reduce the number over the next 3 years, with a permanent ban to take effect from 1996. The majority in the vote was sufficient to override a presidential veto, and Bush signed the bill into law.

The UK position

There is no reference to any of these moves in the UK government's statement on the 1992 Defence Estimates, but simultaneously with its appearance, the Ministry of Defence published a report on the safety of UK nuclear weapons. It was the work of a group set up in 1991 under the chairmanship of the department's Chief Scientific Adviser, Sir Ronald Oxburgh, to consider what bearing the US Drell report had on the safety of the British nuclear arsenal. The names of the members of Oxburgh's committee are not given, but as the report points out, the Atomic Weapons Establishment at Aldermaston is "the sole source of British expertise in the field". The report deals mainly with the institutional arrangements involved in assuring that safety is always a consideration in the design, production and handling of the four operational warheads that Aldermaston has produced or is designing — the free-falling bomb WE177, the original Polaris warhead and its Chevaline modification, and the warheads now under production for the D5 Trident missile which the United Kingdom is buying from the United States.

It is a reassuring document, and its criticisms, such as they are, relate essentially to the division of responsibility in the promotion of safety. Its main recommendation, therefore, is that an "independent champion" for nuclear weapon systems safety should be appointed at the Ministry of Defence, with authority to intervene at any stage in the production and deployment of warheads, and with direct access to ministers — as opposed to reporting through a civil-service hierarchy that may be ill-informed about nuclear matters. (The Drell report also urges organizational reform to ensure identification and proper resolution of safety and surety questions regarding nuclear weapons.) Oxburgh refers to the three modern technical safety criteria discussed both by the Drell panel and by Kidder — a modern detonating system, the use of insensitive high-explosive for the outside casing,

and fire-resistant pits — but does not suggest that the WE177 bomb, which came into service in 1966, should be reconstructed to meet today's safety criteria. According to his report, it will remain in service until at least 2000, when the plan is to replace it with an air-to-surface missile. The Chevaline warhead is due for "retirement" when Trident comes into service, and while Oxburgh does not specifically say so, one must assume that his committee's assessment of the safety of the UK warhead for the D5 missile will have taken into account the point made in US reports about the proximity of warheads to the detonatable fuel of the third stage.

The Oxburgh report also implies that, at some point in the future, testing would be called for. But neither it nor the 1992 Estimates gives any indication that the United States is preparing for what the Americans call a restrictive CTB environment. While the simultaneity of the recent US and Russian moves on a moratorium on underground nuclear tests undoubtedly indicates prior consultation between the parties concerned, there has been no indication that the United Kingdom either took part in, or was even being informed about, whatever negotiations may have been going on. The French have also since declared a moratorium on testing. In recent years, the UK government's stock answer to questions in Parliament about nuclear tests has been that it is necessary for them to continue. The United Kingdom has no test-site of its own, but by a longstanding arrangement has been allowed the use of the one in Nevada — with the United States sharing whatever new information the tests revealed.

The two paragraphs about tests in the UK government's 1992 Defence Estimates can only be described as equivocal, beginning as they do with the words "we are taking a leading role in the Conference on Disarmament in Geneva" — from my experience as a one-time witness and observer, not a very exacting task — and going on to say that the purpose of the 21 UK tests that have been carried out since the Partial Test Ban Treaty (PTBT) of 1963 was "to enhance safety and to establish the effectiveness of new generations of weapons" — that is to say, if one disregards two or three which never left the drawing-board, of the WE177, the original Polaris, and its modified Chevaline version, the only three warheads deployed by the United Kingdom over the past 30 years. (Over the same period the United States developed and deployed at least 30 different warheads. Whether the Soviet Union also produced and deployed as many has not yet been disclosed. The total number of underground tests carried out by the United States between 1963 and

1991 was 605, by the Soviet Union 163, and by France 140.) How many, if any, of the 21 UK tests were specifically made to assure safety is not stated. After the WE177 came into service in 1966, there was no testing for 8 years.

The oblique reference in the Estimates to the need for testing, and the British silence about the moratoria on testing declared by other nuclear-weapons states, is disquieting. The process of getting rid of warheads will be in full swing in the United States and the former Soviet Union when the NPT international conference takes place in 1995. This meeting will differ statutorily from the 5-yearly reviews that have taken place on demand since the treaty came into force. In its 25th year, the states party to it, now numbering about 150, have to decide by a simple majority — which must include the votes of the nuclear-weapon states party to the treaty — whether it continues in force indefinitely, or is extended for fixed periods.

The United States, United Kingdom and Soviet Union took the lead in bringing about the treaty in 1968, with France and China, at the time the only other two nuclear-weapon powers, refusing to be signatories (both have recently signed). India, which exploded a nuclear device in 1974, also refused to be a signatory to the NPT, its overt reason being that the three main powers that were promoting the NPT were continuing to enlarge and diversify their own nuclear arsenals — what became called 'vertical proliferation' — at the same time as they wished to prevent 'horizontal' proliferation.

For India, the NPT and a CTB are inextricably linked. Countries like Iraq could sign and ratify the NPT, but secretly try to become a nuclear-weapon power. Not India. Without proceeding to the manufacture of nuclear weapons, it has demonstrated to the world that it has the knowledge and capacity to make a fission bomb. If by the time of the 1995 meeting worldwide pressure has not made all nuclear-weapon states agree to a test moratorium, or to a specified restriction on underground testing, the issue of vertical proliferation is bound, as it always has been at NPT meetings, to be vigorously condemned, with the United Kingdom singled out as the main offender because of its declared intention to continue testing. On the other hand, it is all but inconceivable that the NPT would be voted out of existence.

The clause in the PTBT that relates to amendments also perpetuates the unique position of the three powers by whom the treaty was negotiated. Any amendment can be passed by a simple majority of those party to the treaty, provided that the majority includes the votes of the three states by whom the treaty was

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Looking to a peaceful future — Yeltsin and Bush sign START 2 in Moscow.

negotiated. That is the only card the United Kingdom now has to put it on a nuclear par with the United States and the former Soviet Union.

Macmillan's mediation

At the start of the nuclear arms race, the United States and the Soviet Union, today the adherents to a test moratorium that could lead to a CTB, were deaf to the widespread cry for a ban on all testing. Within 10 years, however, they had had to bend to world pressure and stop testing in the atmosphere and seas because of radioactive fallout. They were resolute in their refusal to surrender the right to test underground. Some prominent Americans were also strongly opposed to this demand — Adlai Stevenson made the suspension of all testing a major issue in his 1956 presidential campaign against Eisenhower. A few Russians, among them Sakharov, also took an anti-establishment view about any continuation of testing.

Of the world leaders in power at the end of the 1950s, only Harold Macmillan, whose prime ministerial authority prevented any open opposition from UK nuclear weapon experts and the military, passionately fought for a ban on all testing. He tried hard to persuade Khrushchev and Eisenhower and, after him, Kennedy, to follow the line he was urging, but succeeded only in so far as a meeting of US, Soviet and UK scientific experts was convened in 1958 in Geneva — with scientists from other countries present as observers — to consider how a ban on underground tests could be monitored by seismological means to provide assurance that no party to a CTB was testing surreptitiously (there were no scientific problems about detecting tests in the atmosphere or seas).

John Cockcroft and William Penney led the UK team, with instructions to achieve an agreement. They failed. While there was general agreement that it would be possible to differentiate and

locate an underground test from a natural disturbance above a 4.75 reading on the Richter scale, the Americans insisted that shock-waves below that level might be ambiguous, and that even the smallest underground test might have military value. If there was to be a CTB, the Russians would therefore have to agree to on-site inspections of any disturbance that the US weapons-seismologists regarded as suspicious. The Russians were as determined not to agree to intrusive inspections. Macmillan visited Khrushchev to resolve the issue, and the two agreed that the Soviet Union would allow three on-site inspections a year. The United States would not hear of it — it insisted on a minimum of 20 on-site inspections, or no CTB.

It was at this moment, the end of 1959, that I joined the Ministry of Defence as Chief Scientific Adviser, reporting directly both to the minister of defence and to the prime minister, who by now had reluctantly concluded that the US military and nuclear establishment wanted to go on testing indefinitely, as he put it, to further “the art of nuclear weapons”. Conscious of the fact that a state of nuclear deterrence already existed between East and West, I wanted to know what basis there was for the NATO doctrine that, were the Warsaw Pact powers to launch an attack on the West, Western nuclear weaponry could compensate for a presumed inferiority in conventional forces. Analyses of NATO war plans convinced me that as both sides possessed nuclear weapons, a European nuclear battlefield would rapidly become a radioactive waste in which civilians and military alike would perish. The use of nuclear weapons made nonsense of the idea of controlled military operations. Macmillan, with his personal experience of the battlefields of the First World War, needed no persuading about the correctness of my conclusions. He also agreed that as a basis for the concept of mutual deterrence,

no prize that could be gained by force of arms was worth the risk of the total destruction of a country's major cities, with populations of a million or more — a one-megaton warhead could be roughly equated to a million killed or wounded. He became more determined than ever to persuade Eisenhower and then Kennedy to agree to a CTB.

Before he handed over the presidency to Kennedy, Eisenhower had, however, declared openly — it is said, to relieve his successor of the responsibility of so doing — that in view of the continuing failure of the Geneva technical talks, the United States might be impelled to resume testing. The Soviets took this as a signal, and in September 1961, without any warning, began a series of powerful atmospheric tests (including one of more than 50 megatons). The United States followed suit, starting with underground tests, but with the intention of moving on to tests in the atmosphere, for which Kennedy requested the use of Christmas Island. But however massive the Soviet programme, Macmillan could not see that it had significantly changed the existing strategic stalemate. Penney, then deputy chairman of the Atomic Energy Authority, and I were sent to Washington to discover whether our opposite numbers honestly believed that the Soviets had gained anything from their new tests that altered the strategic nuclear balance. The answer was no. Macmillan and Lord Home, the Foreign Secretary, then met Kennedy to argue the issue face to face, but Kennedy insisted that the United States had to respond to the Soviet Union's resumption of atmospheric tests. Because of the complicated network of UK/US relations, Macmillan was constrained to grant the president's request. Glenn Seaborg, then the chairman of the US Atomic Energy Commission (AEC), was later to write that the failure to achieve a comprehensive test ban at the time was “a world tragedy of the first magnitude”.

What we in the United Kingdom had not appreciated was that the nuclear-weapon experts of both sides had used the two-and-a-half-year moratorium during which technical talks were in progress to plan and prepare the largest programme of tests ever carried out. The Soviet series began in September 1961, and by the time it ended in mid-1962, 93 atmospheric and 2 underground tests had been made. The United States began underground testing 2 weeks after the Soviet Union started, and by the time their planned programme ended 6 months later, they had carried out 43 atmospheric and 92 underground tests. Kennedy had obviously been unable to withstand the enormous pressure that the AEC and its friends in the Pentagon and Congress were exercising in favour

of breaking the moratorium. John McCone, Seaborg's predecessor as chairman of the AEC, had been openly campaigning to that end, as had several prominent congressmen.

When Khrushchev refused to agree to more than three annual on-site inspections, the idea of concluding a comprehensive ban evaporated. Before this happened, Penney and I, with a team of UK seismologists, were again sent to Washington to explain to our opposite numbers that the United Kingdom was satisfied that a ban on underground tests could be adequately monitored without on-site inspections. We failed to get them to agree. Negotiations for a ban that allowed underground testing were then begun, and a partial test-ban treaty concluded in August 1963. Not long afterwards, the United States resumed testing, carrying out 15 tests in 1963 and 38 in 1964. The Soviet Union resumed in 1964 with six tests.

The next opportunity to reopen technical talks about the seismological monitoring of underground tests arose in 1966 on the initiative of the United Kingdom. The effort was abortive. When Jimmy Carter assumed the presidency in 1976, he placed a permanent comprehensive ban high on his agenda, but after lengthy negotiations with the Soviet Union and the United Kingdom, his initiative came to nothing. During the exchanges, Harold Agnew, then director of Los Alamos, admitted that given "no restrictions on materials, and adequate non-nuclear testing, the [warhead] stockpile could be maintained for a period of at least 10 years" without any nuclear testing. But under pressure from the nuclear establishment and their political and military allies, Carter's hope for a permanent CTB became reduced first to a moratorium of 5 years, and then to one of 3. All the old arguments against CTB resurfaced, including the likelihood of Soviet cheating and the possibility that the smallest underground disturbance might mean that some device of military value had been tested. The negotiations were suspended after Reagan became president in 1981. The line of his administration was that while adequate verification of a CTB was not achievable, even perfect verification would not make a CTB acceptable, and that "so long as US security depended upon nuclear weapons, nuclear testing must continue".

The turn-around

What Reagan could never have imagined was that his 1987 INF agreement with Gorbachev would not only be followed by the end of the cold war, but to a resurgence of moves to institute a ban on all testing. In January 1991, a PTBT amendment conference was held in New York. Six months later, Sweden intro-

duced a draft CTB treaty at Geneva. If past experience is a guide, the main issues that will stimulate controversy when it, or indeed any alternative draft CTB treaty, comes up for serious discussion will concern the establishment and management of a worldwide network of seismic stations to differentiate between natural underground disturbances and explosions. The Swedish text suggests that a global system consisting of some 220 stations in countries party to the NPT, costing about \$150 million, should be able to detect, locate and identify an underground explosion as small as 0.5 kilotons anywhere in the world.

Obtaining agreement to the principle of on-site inspections to verify suspicious or ambiguous events may also be troublesome for some countries, as would agreeing the nature of the central organization whose responsibility it would be to collect and collate seismic data and to organize inspections. Whether a new body is required, or whether responsibility should be placed upon the International Atomic Energy Authority (IAEA) will undoubtedly be the subject of vigorous debate. The IAEA was set up in the mid-1950s, following on President Eisenhower's UN "atoms for peace" address, to oversee the provision and use of fissile material in the exploitation of nuclear power for civil purposes. In 1968 its scope was extended when it also became the monitoring agency for the NPT. But as its difficulties in dealing with Iraq's illicit nuclear establishments have shown, it does not even now have adequate strength to discharge its dual role.

Nevertheless, if a CTB does become a reality, it would be far better if it, rather than a new body, were given the resources to oversee compliance with either a fully comprehensive ban or a treaty which restricted underground testing. If the United Nations, as was intended at its start, or as has now been suggested by the former Soviet Union, is to exercise a comprehensive and worldwide oversight of the exploitation of the atom, the IAEA must be accorded the resources that allow it to act with the authority that the job demands.

Odd man out

The United Kingdom has not so far responded to the recent moves of the United States, Russia and France regarding a moratorium on underground tests. If, say, in 2 or 3 years' time a decision were taken about a CTB, the only card that the United Kingdom could play, were it to go against the majority, is the critical vote it enjoys in any decision to amend the existing PTBT. It would, however, be a valueless move, now that President Bill Clinton has indicated his agreement to the cessa-

tion of US underground tests by 1996 — and so closing down the Nevada test site.

Presumably the main reason for the United Kingdom standing aside at the moment is a possible wish to retest the Trident warhead that Aldermaston is producing, and to test the one that is being designed for a highly problematical air-to-ground missile. In theory, the United Kingdom could, of course, continue with its present plans without further warhead testing. Its nuclear weaponry is there to deter, not to be used in actual fighting. Nor has it any enemies in its nuclear sights, and there is no likelihood of its being threatened in the foreseeable future by a nuclear aggressor. Trident and the European fighter were conceived in the days of East/West confrontation. The United Kingdom already has Trident, but as the future of the European fighter remains financially uncertain, it would be military nonsense and political folly to continue with the development of an air-to-ground nuclear missile system that lacked any operational significance, particularly if this implied the belief that the United States would allow the United Kingdom to carry out underground tests which they themselves had foregone.

The end of the cold war has not made the world a less turbulent place. Apart from the political upheavals and bloodshed that now plague the old communist world, both it and the West are hard put to deal with the distortion that has been caused in manufacturing industry by military interests since the end of the Second World War. The effort to turn swords into ploughshares is not only creating major domestic economic problems: it has also led to an increase of trade in arms which would have been unthinkable 10 or 20 years ago. This is bad enough in the field of conventional arms, with Western countries competing in the search for customers — mainly in the Middle East — not only with the old Soviet Union and China, but between themselves. The possibility that illicit sales could extend either to nuclear materials or technology is alarming, especially in view of the extreme danger that some unsophisticated people would not be inhibited from using nuclear weapons if they had them. This danger would be lessened enormously were a CTB to be in place to strengthen the NPT, and one serviced by a truly effective IAEA. The nuclear enthusiasts who now stand in the way of a CTB, with promises of new destructive or defensive wonders, are the very people who would be needed to man such a monitoring body. They, too, are swords which could be turned into ploughshares. □

Lord Zuckerman is at the University of East Anglia, Norwich NR4 7TJ, UK.

Sonoluminescence in from the dark

The idea that a sufficiently intense sound wave might make a liquid glow with light is almost an exercise in magic; but the facts are verifiable, the understanding lags behind.

Not everything is yet understood, but that unexceptional statement surely applies with as much force to the shadowy phenomenon of sonoluminescence as to the other great puzzles of our times — where are the missing neutrinos from the Sun, for example, or for what phenomenon is the Big Bang a metaphor?

One puzzle with sonoluminescence is that it seems to be a means of making energy run uphill. A coherent beam of sound waves is made to travel through a liquid whereupon, when the conditions are right, flashes of light can be observed; phonons seem to have been converted into more energetic photons, some of them with ultraviolet energy. Another puzzle is that the duration of sonoluminescent flashes of light, which are at the present limits of fast measurement, seems to be faster than typical electronic transitions in atoms and molecules. What on Earth can be going on?

When reports of phenomena excite disbelief, it is comforting to announce that, far from being unexpected, the phenomena were predicted. That seems to be the spirit in which those working in the field delight in referring to a paper by Lord Rayleigh published in 1917 (*Phil. Mag.* **34**, 94; 1917).

Prolific Rayleigh, with his habit of treating every problem in physics as a candidate question for the Cambridge Tripos examinations, set out to calculate the pressure within an incompressible fluid after a "spherical portion of the fluid is suddenly annihilated". He pointed out the connection with the hissing sound from steam bubbles condensing in a nearly boiling kettle and the way in which the efficiency of ships' propellers is reduced by cavitation. His conclusion was that, at some stage during the collapse of a spherical void in which fluid has been annihilated, almost arbitrarily high pressures may be generated near the boundary, but within the fluid. Sonoluminescence as such is not mentioned, but has been inferred.

Now, mercifully, sonoluminescence is well on the way to being a part of physics, and of experimental physics at that. The mechanism remains that sketched by Rayleigh. Bubbles in a liquid that happen to be at a pressure anti-node (which is, by definition, a stagnation-point of the velocity field and thus a place at which bubbles will hang about) of a sound wave will first be driven to expand and then to collapse catastrophically. Flashes of light appear towards the end of the collapse phase, and can be measured with photomultiplier tubes.

Bradley P. Barber and Seth J. Putterman, from the University of California at Los Angeles, seem now to have gone a long way towards the goal of making sonoluminescent bubbles reproducible (*Phys. Rev. Lett.* **69**, 3839; 1992).

The latest development is the direct measurement of the radii of sonoluminescent bubbles oscillating at the centre of a spherical vessel filled with water, within which a pattern of ultrasonic pressure waves is driven by a pair of piezoelectric elements cemented to the outer walls. (In the measurements described, the period of the sound waves is 37.7 μ s, corresponding to a frequency of 26.5 Hz.) The trick then is to use a fast photomultiplier to record the time course of the light output from a bubble when it is illuminated by a laser beam.

The signal consists mostly of light scattered from an oscillating bubble, expected to depend directly on the radius (or the square thereof). But the system is sensitive enough to record a brief peak of extra light emission just before light scattering reaches its minimum, taken as a mark of when the collapse is complete. When all this is going on, the walls of the cavity are calculated to be travelling inwards at supersonic speed. Imploding is the word for them.

The flash of light appears soon afterwards, and no more than 10 nanoseconds before the bubble has shrunk to its minimum dimensions. The experiments described suggest that the duration of the flashes is no more than 50 picoseconds. A remarkable feature of their appearance is their constancy in phase — the jitter from one sound-wave cycle to the next is also reckoned to be no greater than 50 picoseconds, meaning that the flashes of light can be used as triggers for the recording equipment. And after they have flashed, bubbles appear to oscillate more narrowly in radius, as do bubbles that do not meet the conditions for light emission. Sonoluminescent bubbles are also smaller than Barber and Putterman had expected — bubbles of 4 μ m are efficient sources of flashlight.

The virtue of this development is that it puts the investigation of the sonoluminescent bubble on a solid and reproducible foundation. More than that, Barber and Putterman have shown that the behaviour of their bubbles can be accounted for by the hydrodynamic equations of motion, at least outside the region in which supersonic implosion takes over. By way of fine tuning, they speculate about the way in which energy may be dissipated from the bubble system

by shock waves during the supersonic implosion and by the radiation of sound energy during the post-collapse bouncing phase.

Unfortunately, none of this does much to show what is really going on in sonoluminescence. Nor, for that matter, does the same authors' observation last year (with Robert Hiller) that the spectrum of the light flashed out once every acoustic cycle from their bubbles spans the whole visible spectrum, from red to blue, and then extends into the ultraviolet as far as the frequency at which water becomes an efficient absorber of radiation. In short, the argument then was, the spectrum could well be the tail of a Planck distribution, in which case the corresponding black-body temperature might be 50,000 K or even 100,000 K.

So does this mean that the concentration of acoustic energy during the collapse of a bubble turns its contents into a plasma for a 50 picosecond instant? That is certainly one possibility.

Another, canvassed in 1991 (*Nature* **352**, 318; 1991) when the same authors described their first observations that sonoluminescent flashes keep step with the acoustic fluctuations, is that the mechanism of the flashes may involve some kind of collective emission, as in a laser. All of us will be glad to know how this tale comes out.

Meanwhile, it is as well to keep in mind that sonoluminescence may take different forms. Somehow, collapsing cavities concentrate mechanical energy on a few molecules or atoms in such a way as to evoke unexpected happenings. Everybody seems to agree that the first observation of sonoluminescence was that of the formation of hydrogen peroxide molecules in water nearly 60 years ago. More recently, other groups working with non-aqueous materials have found that the most probable source of light from acoustically assaulted materials is the storage of energy in particular chemical reactions, followed by their luminescent reversal (Suslick, K.S. & Flint, E. B. *Nature* **330**, 553; 1987). Other mechanisms than these are no doubt waiting to be found.

Meanwhile, the rest of us may usefully contemplate a few incidental features of this interesting work. First, Barber and Putterman have repeated on several occasions that they can see the flashes they have recorded with the naked eye. Second, on this occasion the measurements have outstripped the hydrodynamics. This territory is unknown.

John Maddox

Beauty on the brain

Rory Howlett

Is a sense of beauty unique to humans, or is it also a characteristic of other animal species? When considering the evolution of sexually selected male traits such as the tail of the peacock, Darwin was happy to invoke a sense of beauty to explain the apparent preference of females for the gaudiest males. If the study of neural networks has any relevance to what is going on in the brains of female birds when choosing their mates, then the innovative paper by Magnus Enquist and Anthony Arak on page 446 of this issue provides strong, but possibly controversial, support for Darwin's position.

Darwin was pilloried by his contemporaries for his theory of natural selection, which implied that humans had evolved from ape-like ancestors by a process of descent with modification. Equally contentious were his views on the roles of extreme and elaborate male traits in reproduction, as explored in *The Descent of Man, and Selection in Relation to Sex*, published in 1871 (twelve years after the first edition of *The Origin of Species*).

In this book, Darwin set out, among other things, his theory of sexual selection. Realizing that extreme male traits, such as the antlers of deer stags and the plumage of birds of paradise, challenged the view that the morphological and behavioural characteristics of animals could be explained simply in terms of the struggle for existence, Darwin argued that either such traits had evolved because of the mating preferences of females (intersexual selection), or because they had a direct role in the competition between males for access to females (intrasexual selection).

Although the idea that traits such as antlers are involved in male-male competition was and remains relatively uncontroversial, Darwin's views on the role of female choice were hotly debated from the start. Thus, in his book *Darwinism: An Exposition of the Theory of Natural Selection* (Macmillan, London, 1889), Alfred Russel Wallace argued against the notion (explicit in Darwin's writings) that certain female animals, such as peahens, have an aesthetic sense akin to that of humans, a position that Wallace presumably regarded as anthropomorphic.

Part of the problem was that Darwin did not have a satisfactory explanation for the origin and maintenance of the female choosiness assumed to drive the evolution of male traits. His theory was revitalized, however, in 1915, when the great English geneticist R. A. Fisher

presented a plausible verbal model to explain the evolution of female mating preferences — Fisher's so-called 'run-away process'.

Fisher's basic idea was as follows. A novel but essentially arbitrary male trait such as a long tail might appear in a population and enjoy, for whatever reason, a slight selective advantage.

with resistance to parasites, enjoys some support (for a review of these and other theories see M. Kirkpatrick and M. J. Ryan, *Nature* **350**, 33–38; 1991).

Despite the superficial success of theoreticians in providing credible explanations for the coevolution of female choosiness and male splendour, there remains something deeply mysterious about the evolution of traits as exquisite and complex as the eyed tail feathers of a peacock (see illustration), so specific are they in their form. It is perhaps for this reason that people still search for

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Eye for attraction — close-up of a peacock's tail feather.

Females, it was assumed, would be able to distinguish between males, and mutations that caused females to prefer to mate with novel males would increase in frequency because the male offspring of preferential matings would tend to inherit their father's good looks, and female offspring would tend to inherit their mother's choosiness. Assuming that the level of preference could reach some threshold level (perhaps because of genetic drift or direct selection on the mating preference), this would lead to positive reinforcement, and the preference and male trait would coevolve until checked by natural selection against extreme versions of the male traits that were excessively disadvantageous in terms, for example, of viability.

To explain the exaggeration and elaboration of the preferred male trait, Fisher used the rather vague notion of "level of preference". Nevertheless, his basic ideas have found support from a succession of increasingly sophisticated genetic models and computer simulations. Even so, the competing hypothesis that females might prefer certain male traits because they are indicative of genetic fitness, associated, for example,

more satisfying solutions to the puzzle.

One hypothesis that has attracted considerable attention recently is based on the idea that female choosiness and male traits did not evolve together, as assumed by Fisher. Natural selection is likely to shape the sensory systems of females not only because of their role in mate selection, but because they also have other important functions, such as the recognition of warning signals displayed by poisonous prey and their mimics (to choose a random example). Because of the well-established phenomenon of pleiotropy, by which genes can have several effects on the phenotype, perceptual biases could be wired into the brain for reasons quite unrelated to reproduction. Thus, female mating preferences could evolve as a side product of the otherwise adaptive evolution of sensory systems, and may pre-date the occurrence of the preferred male trait. Once established in the female population, such perceptual biases would act as a powerful selective force driving the evolution of male traits.

Although regarded as heretical by neofisherians, the potential importance of such sensory biases in sexual selection

is supported by some empirical studies (see for example M. J. Ryan *et al.*, *Nature* **343**, 66–67; 1990). A major criticism, however, is that these biases can be invoked in an *ad hoc* way to ‘explain’ away the preferences of females for all sorts of weird and wonderful male traits. Without a better understanding of the neuropsychological mechanisms underlying female choosiness, progress in this area is likely to be slow and more in the realm of philosophy than of rigorous science. This is where the study of neural networks as simple, tractable models of female sensory systems may help out.

Enquist and Arak have used artificial neural networks to show that the responses of signal recognition mechanisms are biased, and that signal form can be subject to considerable receiver-driven selection. They first used a procedure mimicking natural selection to train a simple ‘female’ network to distinguish between random patterns and images representing ‘conspecific’ males with long tails and ‘heterospecific’ males with short tails. The network rapidly evolved near error-free recognition by a variety of different routes, demonstrating that there were multiple solutions to the task presented.

Surprisingly, some novel stimuli that had not been used during the training phase but which resembled exaggerated or elaborate versions of the conspecific male elicited stronger responses than the original conspecific stimulus; that is, they acted very much like the super-normal stimuli of classical ethology. Similarly, images of males without tails elicited even weaker responses than heterospecific males with short tails. Enquist and Arak argue that this ability of simple recognition mechanisms to generalize from familiar stimuli to similar but otherwise novel stimuli could exert considerable selective pressure on signal form. This was confirmed by evolutionary simulations in which both male tail length and the female sensory system were allowed to change by mutation, leading to a shift in sensory responsiveness and an exaggeration of the ‘preferred’ trait.

There is no reason, in principle, why Enquist and Arak’s conclusions could not apply equally well to female preferences for any kind of male trait — visual, tactile, auditory or olfactory — although, of course, neural networks may not faithfully mimic the sensory systems of real female birds. It may be anthropomorphic to propose that some animals have an aesthetic sense mechanistically similar to that of humans, but it would be anthropocentric to assume that they necessarily do not. □

Rory Howlett is an associate editor of *Nature*.

Worlds in collision

Gerry Gilmore

DISK galaxies keep their size by balancing the inward force of gravity with rotation. Most of the more picturesque aspects of galactic structure follow from failures of this strict balance. The most common failures lead to spiral structure, bars and such like, where deviations from perfect symmetry collect gas and stars to form regions of high surface brightness, and often regions of in-

branch of modern astronomy. Symmetry in such systems is rare, no doubt as a selection effect as it is the asymmetry which triggers the star formation and our interest in them.

If two galaxies collide nearly head-on, however, there is no force to destroy symmetry and a quite different result occurs — a ring. The principle of such systems is simple, and was explained

long ago by the master of such mergers, A. Toomre (R. Lynds and A. Toomre *Astrophys. J.* **209**, 382–388; 1976). If a compact galaxy passes nearly through the middle of a large disk, then the disk responds just as gravity dictates. While the intruder is in its middle, the disk is forced inwards by the extra mass. Before it comes to equilibrium, however, the intruder departs, leaving the system to bounce back. In the absence of any damping force, the bounce will overshoot, generating an outward-propagating density wave. But nothing has disturbed the radial symmetry, so that the galaxy will look

Wheels of fortune — the Cartwheel galaxy is the unlikely product of a collision between two galaxies.

like a bath after the soap is dropped. A ring galaxy will ensue. A rather similar situation when the target is a large elliptical galaxy has been suggested as the origin of galactic shells. Do such things happen? Indeed, but rarely. It is easy to calculate that only once in a few hundred times will a pair of interacting galaxies be aligned so that one passes nearly face-on through the middle of the other. Actually, this is just about the fraction of interacting galaxies which are ring systems, a statistic which may or may not convey much information. In any case, a few such systems are known where the primary target galaxy now forms a ring, and the projectile galaxy can be identified nearby; the Cartwheel is one of the more spectacular. As its name suggests, this has ‘spokes’, roughly radial features connecting a bright centre to a bright outer ring (see figure).

One other remarkable feature of the Cartwheel is the very low fraction of the chemical elements heavier than helium (‘metals’ in astronomical argot) seen in its outer parts. Such heavy elements are created in supernovae, so their absence suggests a relative absence of supernovae, and hence a relative absence of stars. That is, the Cartwheel, at least

like a bath after the soap is dropped. A ring galaxy will ensue. A rather similar situation when the target is a large elliptical galaxy has been suggested as the origin of galactic shells.

Do such things happen? Indeed, but rarely. It is easy to calculate that only once in a few hundred times will a pair of interacting galaxies be aligned so that one passes nearly face-on through the middle of the other. Actually, this is just about the fraction of interacting galaxies which are ring systems, a statistic which may or may not convey much information. In any case, a few such systems are known where the primary target galaxy now forms a ring, and the projectile galaxy can be identified nearby; the Cartwheel is one of the more spectacular. As its name suggests, this has ‘spokes’, roughly radial features connecting a bright centre to a bright outer ring (see figure).

One other remarkable feature of the Cartwheel is the very low fraction of the chemical elements heavier than helium (‘metals’ in astronomical argot) seen in its outer parts. Such heavy elements are created in supernovae, so their absence suggests a relative absence of supernovae, and hence a relative absence of stars. That is, the Cartwheel, at least

Royal Observatory Edinburgh

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

before it was so rudely perturbed, was a primarily gaseous system which had resisted the temptation to form many stars. This circumstance provides an excellent opportunity to study star formation in a disk relatively free from the complications of disentangling the effects of an old stellar population. This is what is attempted by Marcum *et al.* using new near-infrared imaging from the UK infrared telescope UKIRT in Hawaii.

When a very gaseous ring galaxy is formed, the density wave will behave like a snow plough as it bounces out, collecting up interstellar material. At some stage, as surely as enough rabbits will eventually lead to rabbit pie, enough gas will lead to star formation. Simple models suggest that it is easier to make stars when large masses of gas are squeezed into close proximity than otherwise, so that one might expect star formation to be accelerated at the leading edge of the propagating density ring. If so, a radial gradient in the mean stellar age and in the current star formation rate will be seen. Standard models which predict the colour and luminosity evolution of stellar systems as they age are available, having been developed to analyse data on potentially young galaxies at very high redshift.

The Cartwheel, being simple, big, close and manifestly young, is a perfect place to test such models. And in crude terms it agrees with them. The star formation rate is certainly greatest at the leading edge of the ring at present. Also, Marcum *et al.* find an exceptionally steep radial gradient in the colour of the Cartwheel, not inconsistent with the predictions of most models, though the detailed agreement is far from encouraging for those who interpret data at high redshifts.

Such quibbles aside, however, the Cartwheel does indeed nicely fit into the picture outlined above. A gas-rich disk galaxy has been shocked by a compact intruder, which passed at near-normal incidence near its middle. The resulting gravitational bounce, and no doubt a bit of splash as well, generate an outwardly propagating density wave, which sweeps up the disk gas like a snow plough. In the highest density regions star formation is induced, though it ceases after the density peak passes. The resulting burst of star formation illuminates the ring, bringing the system to our attention, and generates a substantial amount of chemical enrichment of the remaining gas. The 'spokes' are plausibly explained as the remnants of unusually long-lived regions of star formation, which persisted after the main density wave passed. High-precision studies of this galaxy now provide an ideal opportunity to test current models of the evolution of young star-forming systems, and so to

improve our understanding of the environmental dependence of the rate of star formation. It will also be a valuable test bed for quantitative testing of those models of the luminosity and colour evolution of young galaxies which are used to interpret data on galaxies at very high redshifts.

There is one other question. The Cartwheel managed to survive as a large

gas-rich disk galaxy until recently with almost no star formation. How? And as we notice it now only by the extremely unlikely chance of a face-on collision, how many other such 'primordial' disks are out there waiting to be found? □

Gerry Gilmore is at the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK.

INSECT DEVELOPMENT

The long and the short of it

Vernon French

How general are the mechanisms of spatial patterning in early insect embryos? Two papers, one published in May of last year¹, the other on page 448 of this issue², reach rather different conclusions, especially as regards the function of the pair-rule genes whose striped expression is the first sign of periodicity during the process of segmentation in *Drosophila*.

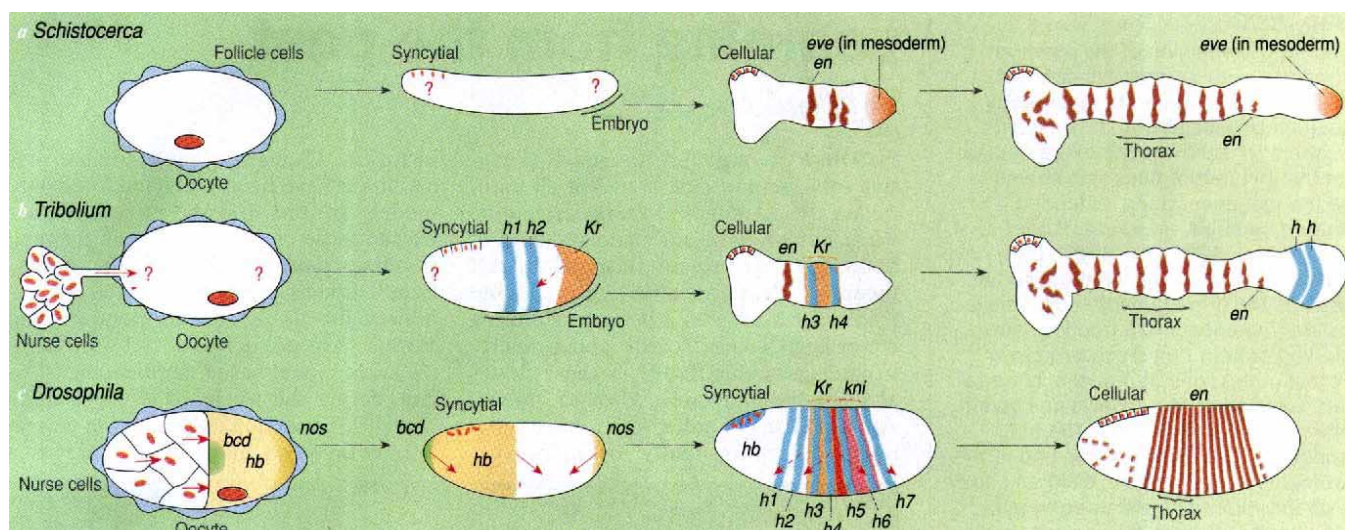
Early insect development is peculiar in that many nuclear divisions take place within the peripheral cytoplasm, forming a syncytial layer from which the cellular blastoderm arises. In so-called long-germ insects (such as *Drosophila*), the embryo develops from most of the blastoderm, and divides up more-or-less synchronously into segments³. By contrast, in short-germ insects, the small early embryo appears to comprise only the prospective anterior head and a growth zone, the remaining segments being formed sequentially during posterior cell division. Other species are termed intermediate-germ, as only the abdominal segments appear sequentially during growth of the posterior end of the embryo³.

Molecular genetics has brought spectacular revelations about the mechanism of segmentation in one long-germ insect, *Drosophila* (see figure). The story runs like this. Specific maternal transcripts (such as *bicoid* and *nanos*) are imported from adjacent nurse cells, localized at anterior and posterior ends of the oocyte, and translated in the early embryo to give long-range protein gradients in the syncytial cytoplasm. High levels of *bicoid* protein activate zygotic expression of the anterior gap gene, *hunchback*, and *nanos* prevents posterior translation from the maternal *hunchback* transcripts which entered the egg during oogenesis⁴. The resulting gradient of *hunchback* protein activates bands of expression of other zygotic gap genes such as *Krüppel* and *knirps*⁵. From the sites of production, the gap proteins diffuse in the still-syncytial cytoplasm to form overlapping gradients. Particular

combinations of gap-protein concentrations then activate pair-rule genes such as *hairy*⁶ and *even-skipped* (*eve*)⁷, resulting in the characteristic seven-striped expression pattern over much of the embryo. By this stage, the blastoderm nuclei are partially isolated by membranes, and pair-rule proteins do not diffuse away. The sets of stripes are out of phase, and particular pair-rule combinations activate segment-polarity genes such as *engrailed* and *wingless*⁸, establishing a segmentally striped expression pattern.

In the long-germ embryo of the fly, *Musca*, the expression patterns of *bicoid* and several zygotic segmentation genes resemble those in *Drosophila* (although with some intriguing differences), indicating that a similar mechanism is involved⁹. But what of other types of insect embryo? Visible segmentation is preceded by the sequential appearance of stripes of *engrailed* expression in the short-germ embryo of the locust (*Schistocerca*)¹⁰. It has been argued, however, that the earlier stages must differ dramatically from those in *Drosophila*, because they occur in a cellular epithelium within which diffusion gradients of large transcription factors (like gap proteins) could not form¹⁰. The pair-rule genes, particularly, might be specific to long-germ insects^{1,10,11}. Even in *Drosophila*, *engrailed* expression is controlled after the blastoderm stage by cellular interactions mediated by segment-polarity genes such as *wingless*, and the pair-rule genes do not function at all in the anterior head segments. The idea that insect segmentation mechanisms are diverse was supported by the demonstration¹ that the locust expresses a homologue (or 'orthologue') of *eve* in the nervous system (as does *Drosophila*), but not in a pair-rule pattern during segmentation.

Now, however, Sommer and Tautz² show that the 'short-germ' flour beetle (*Tribolium*) does express orthologues of the gap gene, *Krüppel*, and of the pair-rule gene, *hairy*, in patterns resembling



Simplified representation of oogenesis and segmentation in *a*, the locust *Schistocerca* (panoistic ovary, short-germ embryo); *b*, the beetle *Tribolium* (meroistic ovary, short- or intermediate-germ embryo); and *c*, the fruitfly *Drosophila* (meroistic ovary, long-germ embryo). Colours indicate the location, where known, of transcripts from the maternal genes *bicoid* (*bcd*) and *nanos* (*nos*); the gap genes *hunchback* (*hb*), *Krüppel* (*Kr*) and *knirps* (*kni*); the pair-rule genes *hairy* (*h*) and, in *a*, *even-skipped* (*eve*); and the segment-polarity gene, *engrailed* (*en*). The dashed arrows indicate diffusion to form a protein gradient in the syncytium.

those of *Drosophila* (see figure). Two transient stripes of *hairy* expression appear in the blastoderm, and there is also a more posterior band of *Krüppel* expression (apparently corresponding to the future thorax) within which two transient *hairy* stripes subsequently arise, as in *Drosophila*⁶. During embryo extension, two further *hairy* stripes are seen in the growth zone and these, too, may be transient (the anterior one fading as cells leave the zone, while another forms in new posterior cells). In all regions of the embryo, *engrailed* stripes follow those of *hairy*, but the precise relationship between the two has not been established.

These new findings suggest, intriguingly, that the mechanism found in *Drosophila* does operate in the segmentation of short-germ embryos². *Tribolium*, however, is normally classed as 'intermediate-germ', with only the abdominal segments forming by posterior growth (although it is difficult to make this distinction without cell labelling and reliable early fate maps³). Perhaps, in *Tribolium*, gradients of anterior gap proteins (such as *Krüppel*) form in the early syncytial stage and can then control *hairy* stripes 1–4 in the same (poorly understood) way that they do in *Drosophila*⁶, but in a cellular system. The abdomen, however, must be rather different, because its *hairy* stripes appear during growth of an epithelium, where control by diffusion is unlikely. Perhaps a pulse of expression of a posterior gap gene (such as *knirps*) could result in a graded protein distribution through the dilution effect of posterior cell division — further study of this system should prove fruitful.

Regardless of doubts concerning the

'shortness' of *Tribolium*, *hairy*¹ and *eve*² do seem to differ with respect to epithelial segmentation. Sommer and Tautz² resolve this problem by suggesting that the pair-rule stage is indeed fundamental to insect segmentation, but that only *hairy* is part of the ancestral mechanism. In *Drosophila*, however, both genes are regulated directly by gap proteins^{6,7}, and *eve* (but not *hairy*) seems to regulate directly the expression of *engrailed*⁸. Also, there are reasons to suspect that the difference between the two studies derives from the insects, not from the genes. *Schistocerca* (like all orthopterans) has the ancestral (panoistic) form of oogenesis, with no nurse cells. *Tribolium* (like other beetles, and dipterans such as *Drosophila*) has meroistic oogenesis, with the oocyte provisioned by anterior nurse cells. As all zoologists know, there are an awful lot of beetles. Their embryos are also very diverse, ranging from short- to long-germ, whereas it seems that insect orders with panoistic oogenesis have not evolved long-germ segmentation³. These forms of insect diverged some 300 million years ago, and it is likely that meroistic oogenesis enabled the evolution of new ways of establishing the embryonic axis and patterning it rapidly, during the early syncytial phase of development^{3,11}. Short-germ embryos derived from the two forms of insect oogenesis may indeed have different mechanisms (and different types of segmentation gene) for establishing the pattern of *engrailed* expression, which is a conserved feature of arthropod segmentation.

This issue could be rapidly resolved. We may soon read whether locust *hairy* or beetle *eve* is expressed in stripes and, therefore, whether pair-rule genes are

indeed likely to be fundamental to segmentation. Also, orthologues of *hunchback* have been isolated from both of these insects¹², and expression patterns should emerge soon, giving clues about the function and regulation of gap genes. In *Drosophila*, localized expression of the homeotic genes is established around the cellular blastoderm stage, conferring segmental identity, and orthologues of many of these genes, too, are now known in beetles¹³ and locusts¹⁴.

The polymerase chain reaction and *in situ* hybridization on a wide range of arthropod and annelid embryos can find the players (those known from *Drosophila*, anyway) and reveal their locations during segmentation in all its forms (long-germ, short-germ, and with and without determinate lineages). To go beyond speculation, however, we will eventually need methods of directly testing what they are doing there. □

Vernon French is in the Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK.

1. Patel, N. H., Ball, E. E. & Goodman, C. S. *Nature* **357**, 339–342 (1992).
2. Sommer, R. J. & Tautz, D. *Nature* **361**, 448–450 (1993).
3. Sander, K. *Adv. Insect Physiol.* **12**, 125–238 (1976).
4. Nüsslein-Volhard, C. *Development Suppl.* **1**, 1–10 (1991).
5. Hülkamp, M. & Tautz, D. *Bioessays* **13**, 261 (1991).
6. Pankratz, M. et al. *Cell* **61**, 309–317 (1990).
7. Stanojevic, D. et al. *Science* **254**, 1385–1387 (1991).
8. Ingham, P. W., Baker, N. E. & Martínez-Arias, A. *Nature* **331**, 73–75 (1988).
9. Sommer, R. & Tautz, D. *Development* **113**, 419–430 (1991).
10. Patel, N. H. et al. *Development* **107**, 201–213 (1989).
11. Akam, M. & Dawes, R. *Curr. Biol.* **2**, 395–398 (1992).
12. Sommer, R. et al. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
13. Stuart, J. J., Brown, S. J., Beeman, R. W. & Denell, R. E. *Nature* **350**, 72–74 (1991).
14. Tear, G. et al. *Development* **110**, 915–925 (1990).

RÉSUMÉ

Cap that

THE vernier precision of the myofibril structure, with its crystalline arrays of thick and thin filaments, must require gauges for assembly and rivets and washers for survival. The length gauge for the thin (actin) filaments seems to be the immensely long, extended protein, nebulin. Now V. M. Fowler *et al.* (*J. Cell Biol.* **120**, 411–420; 1993) have identified what looks like the cap that secures the thin filament ends. This is called tropomodulin, a tropomyosin-binding protein first discovered by Fowler in red cells. It prevents head-to-tail association of tropomyosins bound along the actin filaments. The tropomodulin sits at the free end of the filaments (distal to the Z disk) and so probably maintains the tropomyosin-stabilized filaments at constant length.

True grit

MAKING the best of a bad job, N. Humphrey *et al.* have extracted information about the layer of debris (till) between glacial ice and the bedrock in the Columbia glacier, southeast Alaska (*J. geophys. Res.* **98**, 837–846; 1993). Drilling at a lower site, the authors found the till occurred only as a veneer, 7 cm thick. At a second site, higher up the glacier, what they describe as an “unplanned but unique experiment” occurred. The drill stuck in the rocky debris and when it was hauled free, days later, it had been severely bent by the movement of the glacier. By analysing the deformation, scrapes and gouges, Humphrey *et al.* conclude that the debris was about 65 cm thick at that point and had a plastic yield strength of some 5.5×10^3 Pa, not enough to resist glacier flow. Others who feel that their research is stuck should take heart.

Three from one

By building evolutionary trees and applying statistical geometry in sequence space to them, F. R. Schubert and colleagues delve further into the history of the highly conserved homeobox genes which are involved in the control of development (*Proc. natn. Acad. Sci. U.S.A.* **90**, 143–147; 1993). Schubert *et al.* analysed nucleotide sequences of 38 such genes from humans and 11 from *Drosophila*. The upshot is the claim that Antennapedia-type homeobox genes split into three classes early on in the evolution of metazoans. The authors suggest, moreover, that the progenitor may have been a three-gene cluster (itself derived from a single gene) consisting of the forerunners of the 3', central and 5' classes (which, they point out, correspond to the anterior, central and posterior parts of the embryo). If these striking proposals stand up to scrutiny, the lid of the homeobox will have been prised open yet a little further.

FLOOD HAZARDS

Learning from the past

Vic Baker

FLOODING is arguably the most pervasive, diverse and destructive of all natural hazards. A hundred years of scientific study has brought our understanding of flooding to its present state of abstract theoretical sophistication. But despite the alleged progress, we are suffering increasing damage to life and property from catastrophic floods — those floods at the large and rare end of the range. Are flood magnitudes really increasing, is our perception faulty, or are we be-

historical. Changes in mean annual precipitation have been predicted in climate models of the greenhouse effect that exceed those responsible for the changed flooding patterns reported by Knox. Some models predict increased probabilities for droughts and flood-causing storms. Are politicians to rely primarily on such model-based predictions when considering the threat of global climate change? There are considerable uncertainties involved, and mistakes (of wrong

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Lafayette, Louisiana, on 20 January after 27 cm of rain had fallen in 24 hours. Floods are arguably the most destructive of all natural hazards — are they increasing in magnitude?

coming careless about where we live and work?

On page 430 of this issue, James Knox provides a long-term historical perspective on this problem by reconstructing a 5,000-year record of overbank floods on Mississippi River tributaries in the north-central United States. His detailed analysis of the largest floods as individual events shows that relatively modest climatic changes during past millennia caused comparatively large changes in flood magnitudes.

The largest floods 5,000–3,300 years ago were, by today's standards, relatively unimpressive. But with only slight cooling 3,300 years ago, far larger floods became common — the sort that we would expect to see only once every 500 years. Further changes occurred during the lead up to the Little Ice Age, around 600 years ago. Knox found these trends by identifying cobbles and boulder gravel washed over the tributaries' banks, calculating the flow depths necessary to have carried them over and dating them through interlaced organic matter.

The question isn't merely academic or

action or inaction) could be costly.

One response to global change has been to call for massive observation programmes, involving satellites, aircraft and global networks of ground stations. Besides the drawback of costs, it is not clear whether such programmes can really help. For one thing, we may end up just monitoring the approach of catastrophic global change, but be too late to react. Another strategy is to try to extrapolate from the vast data available from common experience — in the present case, with smaller floods. This works only if the hydrological processes that cause floods are unchanging and if the historic recurrence times reflect the threat now. But Knox's data show this not to be the case. A small change in climate can cause a large change in flood patterns.

Knox also mentions anecdotal evidence from China, spanning 1,000 years and encompassing the Little Ice Age, that seems to bear this conclusion out. Certainly, looking at the largest floods in the past shows that there is no simple proportionality between the scale or frequency of floods and climate variations.

The vast repository of evidence recorded geologically over the millennia is the best place to find these floods.

Knox deliberately chose to study the flood-prone Upper Mississippi, because it drains a large area and because it lies at the confluence of two climate systems, making it particularly sensitive to climate variations. Now that we see that the

approach works there, we should try looking in the tropics where monsoons, storms and El Niño/Southern Oscillation climate fluctuations may be particularly sensitive to greenhouse warming. □

Vic Baker is in the Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA.

MOLECULAR EVOLUTION

Genes-in-pieces revisited

W. Ford Doolittle and Arlin Stoltzfus

ON page 470 of this issue¹, Tittiger *et al.* describe an intron in the triosephosphate isomerase (TIM or TPI) gene of the mosquito *Culex tarsalis*, an intron that neatly subdivides an exon that would otherwise encode two protein 'modules'. This is exciting because it confirms a prediction by Gilbert *et al.*², and may bring joy to supporters of the 'introns early' theory of gene evolution.

First some history. Immediately upon the discovery of introns in 1977, Walter Gilbert proposed in these pages³ that the split gene organization of eukaryotes serves evolution. By separating protein-coding information into many little bits, he ventured, introns facilitate genetic rearrangements which offer up novel mosaic genes to the scrutiny of natural selection. For some of the recently assembled genes of the vertebrate blood-clotting system, Gilbert has been proved right — mosaic evolution respecting exon boundaries is observed⁴.

A bit later, Darnell⁵ and one of us⁶, reasoning that the long-term advantage conferred by the ability to shuffle exons could not itself account for the existence of introns, ventured that these intervening sequences were really evolutionary relics. In this 'introns early' view, spliceosomal introns mark the positions at which ancient (and independently replicating) exons were joined to form the first modern-sized genes. Their absence from present-day prokaryotes could be due to sudden or gradual deletion over the past three billion years — 'streamlining'.

Another way to explain split genes ('introns late') has it that spliceosomal introns are descendants of transposable elements which invaded the eukaryotic lineage sometime after its divergence from prokaryotes. Some spliceosomal introns do indeed seem to have arisen recently and some introns do contain mobility-conferring genes⁷. However, proving each and every contemporary intron a relic is not the point. The real question is: are at least some modern exons the descendants of small, separately evolved units from which the

first genes were assembled?

Two different sorts of data seem relevant. The first, surveys of the phylogenetic distribution of individual introns and the spliceosomal machinery, are not friendly to 'introns early'. Introns in similar positions of homologous genes (TPI genes, for instance) from animals, plants and fungi are almost irrelevant. There are very many independent evolutionary divergences in eukaryotes much deeper than these, and still many below that leading to euglenoids, the most deeply diverged lineage known to have introns⁸. Lineages further from us than euglenoids may lack introns and the spliceosomal machinery. If so, we must imagine many independent instances in which eukaryotic (as well as prokaryotic) genomes would have had to be streamlined — possible, but very unparsimonious^{7,9}. (Group I introns found in eubacteria have no bearing on the situation, in the absence of any other reason to believe them ancestral to spliceosomal introns.)

The second sort of relevant data harkens back to Blake's 1978 suggestion¹⁰ that jumbling protein pieces together at random, without regard to structural motifs and functions, would produce only garbage. Blake's original view that exons should encode domains has been revised, and now advocates of the introns-early theory seek a variety of correlations at lower levels of protein structure¹¹. Indeed, believable correlations between gene structure and protein structure can be seen for some genes recently produced by shuffling — like those of the blood-clotting system⁴.

But again, this limited data set says nothing about the ancestral situation. Repeated shuffling selects for exons that mix well together, even if they were created initially by insertion of intron/transposons at positions that were random with respect to protein structure. So Gilbert *et al.*² proposed that we focus on ancient genes, produced before the earliest evolutionary divergence (that between eubacteria and eukaryotes). If such genes boast exons that encode dis-

crete units of structure or function at the protein level, those exons must have had previous evolutionary history as discrete units, possibly even before the first cells.

TPI has been one of the three or four favourite ancient proteins. Gilbert *et al.*², after examining the structures of TPI genes from chicken, maize and *Aspergillus*, deduced that the ancestral form had eleven exons, each encoding a discrete unit of structural compactness, as defined by Go¹¹. Nine of the original ten introns were found in at least one modern gene. The missing tenth intron would have divided the third modern exon in two, and Gilbert *et al.* ventured to "predict that in some other [TPI] gene there will be an intron that breaks up this exon". It is this prediction that Tittiger *et al.* have made real.

So does this save the day for introns early? It depends on how sympathetically one reads ref. 2. The null expectation of failure was as low as zero (as forbidden intron positions were not specified), or as high as 92 per cent (as the context suggests that nearly all possible positions can be viewed as forbidden — a strict interpretation implying that, in future, no new intron positions are expected). These odds are more compelling than the previous detailed analysis of TPI by Go and Nosaka¹²; with the broad definition of module boundaries allowed by these authors, most randomly placed introns would have been taken as consistent with introns early.

Prudence calls for a more rigorous analysis of many more of the data. Ideas about what are the relevant features of protein structure can be too easily modified to fit the gene structure data better. Analytical rules that have evolved in the analysis of TPI should therefore be applied blindly to all other proteins for which there are adequate structural and genetic data. To our knowledge, the essential step of excluding the null hypothesis (of randomly placed introns) has not been taken, although there are surely now enough data to attempt this.

W. Ford Doolittle and Arlin Stoltzfus are in the Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada.

1. Tittiger, C., Whyard, S. & Walker, V. K. *Nature* **361**, 470–472 (1993).
2. Gilbert, W. *et al. Cell* **46**, 151–154 (1986).
3. Gilbert, W. *Nature* **271**, 501 (1978).
4. Blake, C. C. F., Harlos, K. & Holland, S. K. *Cold Spring Harbor Symp. quant. Biol.* **52**, 925–931 (1987).
5. Darnell, J. E., Jr *Science* **202**, 1257–1260 (1978).
6. Doolittle, W. F. *Nature* **272**, 581–582 (1978).
7. Cavalier-Smith, T. *Trends Genet.* **7**, 145–149 (1991).
8. Muchal, U. S. & Schwartzbach, S. D. *Pl. molec. Biol.* **18**, 287–299 (1992).
9. Palmer, J. D. & Logsdon, J. M., Jr *Curr. Opin. Genet. Dev.* **1**, 470–477 (1991).
10. Blake, C. C. F. *Nature* **273**, 267–268 (1978).
11. Go, M. *Nature* **291**, 90–92 (1981).
12. Go, M. & Nosaka, M. *Cold Spring Harbor Symp. quant. Biol.* **52**, 915–924 (1987).

Flickers within cycles

Scott Lehman

CRUDELY, the Earth's climate seems to come in two states — glacial and interglacial — that persist for thousands to tens of thousands of years. But signs of much shorter-period variations are crowned in this issue¹, with the report of climate 'flickers' lasting as little as 10 years or less throughout much of the past 40,000 years.

The results, reported by Taylor *et al.* on page 432, are the first data to come from the GISP2 (Greenland Ice Sheet Project 2) group which has been drilling at Summit on the Greenland Ice Sheet for the past four years. The GISP2 core is one of two being recovered as part of an international effort to monitor a variety of climate proxies and atmospheric gases locked away within the ancient ice. The location (at 3,230 metres) was chosen, as an improvement on earlier drill sites, to provide greater stratigraphic resolution and length of record, fewer disturbances due to ice flow, and to obtain ice less likely to have been affected by summer melting — an important source of uncertainty in CO₂ measurements. A key reason for studying two cores in parallel is to aid in distinguishing the common climate signal from glaciological and analytical noise, an increasingly difficult distinction as quantities within the ice are measured on the seasonal-to-annual timescales resolved by the new cores.

The European consortium (GRIP) presented its first results last September². Their ¹⁸O measurements established that the climate was bistable during much of the last ice age, showing that every 500 to a few thousand years air temperatures shifted by an average of 6 °C between mild and cold episodes, each switch taking about 50 years. Similar large and abrupt temperature oscillations have been seen from the end of the last ice age, and the GRIP variations were probably driven by the same mechanism — changes in the mode of circulation in the North Atlantic Ocean^{3,4}.

Taylor *et al.* have now fulfilled one important objective of the two-core strategy — their isotopic and layer-counting results confirm the amplitude and absolute timing of abrupt century-to-millennial scale climate changes during the last deglaciation and also the bistable nature of glacial climate. Where their report breaks new ground is in demonstrating that the longer-term climate shifts seen in earlier records are themselves marked by strong variations of less than ten years duration.

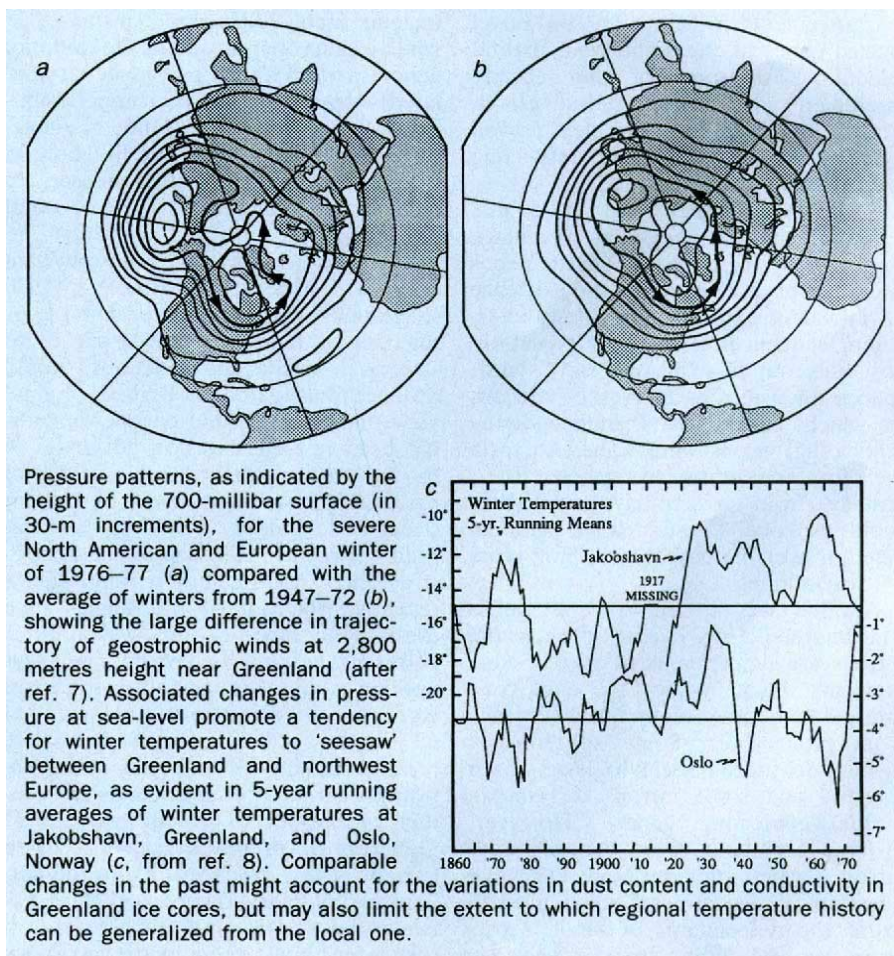
Their evidence comes from measure-

ments of the core's electrical conductivity, a quantity that responds primarily to changing dust concentrations in the ice. Using 15 samples or more a year, Taylor *et al.* obtained a record of dustiness for the past 40,000 years with seasonal-to-annual resolution. Although the link between dust and climate is not fully understood, there is a clear relationship between cold climate and increased dust loading of the atmosphere (and of the ice) on timescales of millennia to centuries⁵, most probably due to changes in the extent of source regions south of the former ice sheet in North America and air-mass turbulence and routing over the northern Atlantic. The sensitivity of conductivity measurements to changes in actual dust concentration within the ice appears to be greatest during intervals of mild climate: during cold intervals the generally high dust loadings saturate the ice and further changes in dust content become undetectable. Despite this lack of sensitivity at the cold end of the climate spectrum, the records provide valuable insight into climate variability during periods of

overall warmth. Most remarkable is the tendency of conductivity to reveal pronounced decade-long oscillations as climate shifted between longer-term modes.

For example, during the well-studied transition from the Younger Dryas cold period to the Holocene (the present interglacial), when average temperatures over the ice sheet rose by 7 °C in 50 years⁶, the conductivity underwent three distinct order-of-magnitude oscillations (equating to fourfold changes in actual dust concentration⁶). Individual transitions spanned no more than five years each. A similar pattern is repeated at the beginning and end of each of the longer-term climate shifts of the last deglaciation. As it is improbable that the location or extent of dust sources in North America could change greatly on such short timescales, these oscillations are much more likely to be rooted in changes in air-mass trajectory or changes in soil moisture at the source region.

Taylor *et al.* refer to the oscillations at major climate transitions as flickers, implying that large-scale mode switches of the ocean-atmosphere system may have been briefly indecisive. Although pronounced decade-long oscillations of Atlantic circulation might occur as it settles into a new mode, the conductivity



record also displays marked variability throughout the mild parts of the ice age and its termination, and strong, persistent short-term oscillations of Atlantic circulation at these times seems unlikely (although some ocean models do produce such behaviour). Inasmuch as conductivity measurements indicate qualitatively similar variability during the early part of the Holocene and preceding mild intervals, it would appear worthwhile to look to the historical record of climate for evidence of interannual and decadal-long oscillations between different preferred states of atmospheric and oceanic circulation.

One example that might produce changes in dustiness over Greenland is the North Atlantic Oscillation — a well-known tendency for negative correlation between severity of winters in Greenland and northwest Europe, and between sea-level pressure in the vicinity of Iceland and along a belt from the US east coast to the Mediterranean⁷. In addition to producing a dramatic 'seesaw' in winter temperatures between Greenland and northwest Europe⁸, it is also marked by a significant shift in the orientation of winds aloft (see figure). Some of the most notable changes in wind-steering occur near Greenland. During the particularly severe winter in North America of 1976–77, when the seesaw tilted in the extreme, the pattern of pressure aloft was similar to the normal one but greatly exaggerated⁹. In addition to the obvious virtue of simplicity, this attribute of the North Atlantic Oscillation may have further attractions, because it suggests that this modern flicker could be amplitude-modulated in response to longer-term changes in boundary conditions.

Whatever the cause of the high-frequency changes at Summit, the results of Taylor *et al.* show clearly the capacity of the new ice cores to provide records of climate comparable to the instrumental record, but with the potential of spanning the last 200,000 years. At the bottom line, the interest is in what these may teach us about possible greenhouse warming. On the one hand, changes such as the North Atlantic Oscillation show that we will have to be very careful about generalizing from the

local changes in Greenland. Longer-term variations reinforce this message: although the global mean temperature has risen by about 0.5 °C in the past few decades, temperatures around the North Atlantic and Greenland have fallen by the same amount¹⁰. Average winter temperatures on the coast of Greenland have repeatedly changed by up to 8 °C in the past century, putting the 7 °C change in mean temperature over the 50-year close of the Younger Dryas into a different perspective. Deciphering the new ice cores may not be easy: we can expect them to portray a complex response to

climate change on a variety of timescales, from millennia to years, forced by a variety of local-, regional- and global-scale factors.

The geological record of such a diversity of variations in the climate poses climatologists and policy makers with an equally challenging puzzle — how long will we have to wait before we are convinced that we are seeing or not seeing a greenhouse signal? □

Scott Lehman is at Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

PARKINSON'S DISEASE

Fetal implants put to the test

Fred H. Gage

IMPLANTING human fetal tissue into humans with neurodegenerative diseases is fraught with biological, clinical and ethical problems, but if the humans who receive the implants show significant improvement the problems may be worth addressing. Three articles in the *New England Journal of Medicine*^{1–3} force a closer look at whether human fetal-tissue implants of mesencephalic cells to the basal ganglia of patients with Parkinson's disease or drug-induced parkinsonism can be called a successful experimental therapy. This issue takes on added significance in the United States, given that President Clinton has just lifted the ban on government sponsorship of research involving the use of fetal tissue.

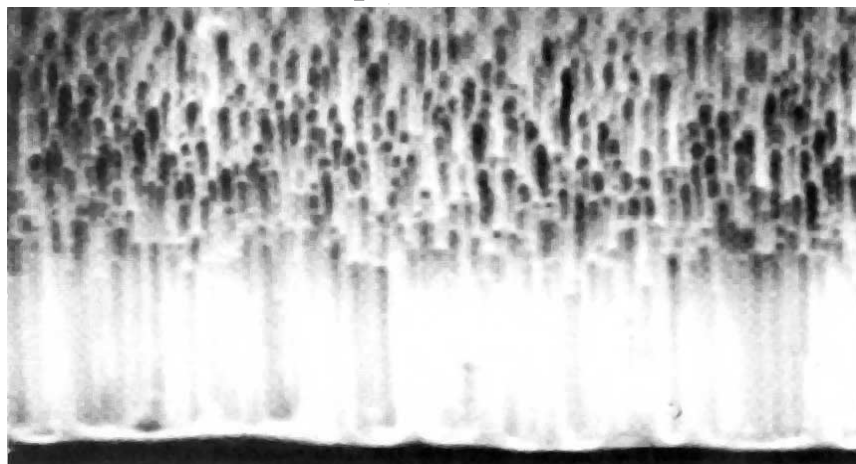
Fourteen years ago, reports on the first fetal-tissue implants in rat models of Parkinson's disease documented anatomical and functional repair^{4,5}. The reliability and specificity of this procedure are adequately documented in rodents, and provide the scientific rationale for human tests. The leap from rats to humans was made first with chromaffin cells from the patient's own adrenal glands, which contain catecholamines similar to those found in the dopamine cells in the brain. Although the initial clinical use of these autologous peripheral cells was reported to be successful, lack of replication by many investigators resulted in justified scepticism. But with human tests of peripheral-cell transplantation underway, it was a short but important jump to trying human fetal-tissue implants in humans. Several reports in 1990^{6,7} indicated that some level of recovery in Parkinson's disease was possible using human fetal mesencephalic tissue, but the degree of recovery was not therapeutic. These initial reports led to more extensive tests on a larger number of patients, culminating in the three new reports.

Despite the work of three separate research teams, a total of 13 patients, similar outcome measures, and a common therapeutic strategy, generalizations cannot easily be made; the differences in experimental design are striking and make comparisons between the studies nearly impossible. However, using common objective clinical measures, all three studies report significant improvement of motor performance in some of their patients. Of the two studies on patients with Parkinson's disease, Freed *et al.*² report greater success than Spencer *et al.*¹ The most encouraging results are reported by Widner *et al.*³ with two cases of drug-induced parkinsonism (see especially their PET scans in ref. 3). Some basic methodological issues previously addressed in preclinical studies on rodents and monkeys may shed light on the reason for the relative successes and provide recommendations for the design of future human experiments.

Two of the main questions to be resolved are: (1) what tissue should be used, and (2) where should it be put? The type of tissue has generally been agreed upon: fetal mesencephalon, which is rich in the developing dopaminergic neurons that are lost in Parkinson's disease. But that is where the similarities in these studies end. Spencer *et al.* used one 7-week-old, two 10-week-old and one 11-week-old fetuses, and froze them in liquid nitrogen for up to 10 months before implantation. This protocol allowed them to test the tissue for viruses, fungi and bacteria. Freed *et al.* used 6- to 8-week-old fetal tissue, keeping the cells in culture for up to seven days to determine the quality of the tissue before implantation. Widner *et al.* also used 6- to 8-week-old fetal tissue, but implanted the tissue within 4 hours of the abortion. Tissue of a gestational age of more than 10 weeks may be more

1. Taylor, K. C. *et al.* *Nature* **361**, 432–436 (1993).
2. Johnsen, S. J. *et al.* *Nature* **359**, 311–313 (1992).
3. Broecker, W. S. *et al.* *Nature* **315**, 21–25 (1985).
4. Lehman, S. J. & Keigwin, L. D. *Nature* **356**, 757–762 (1992).
5. Hammer, C. U. *et al.* *Greenland Ice Core: Geophysics, Geochemistry, and the Environment* (eds Langway, C. C., Jr, Oeschger, H. & Dansgaard, W.) 90–94 (American Geophysical Union, 1985).
6. Dansgaard, W., White, J. W. C. & Johnson, S. J. *Nature* **339**, 532–534 (1989).
7. Waller, G. T. & Bliss, E. W. *World Weath. V. Mem. R. Met. Soc.* **4**, 53–84 (1932).
8. Namias, J. *Mon. Weath. Rev.* **106**, 279–295 (1978).
9. van Loon, H. & Rogers, J. C. *Mon. Weath. Rev.* **106**, 296–310 (1978).
10. Jones, P. D. *et al.* *Clim. Monit.* **16**, 175–185 (1988).

Bristling with potential



You've heard of silicon valley, now meet semiconductor forest. The micrograph shows a dense pattern of indium arsenide whiskers — over 10^{10} of them per square centimetre, each 20–30 nanometres in diameter. They were prepared by Masamitsu Yazawa and colleagues at Hitachi's Central Research Laboratory (*Appl. Phys. Lett.* 61, 2051–2053; 1992). The researchers are interested in the quantum-mechanical effects that arise in fine semiconductor wires, in which electrons can move in one dimension only — along the wire and not across it. Quantum wires are often etched out of a film of semiconductor. In previous work, Yazawa and colleagues have grown the whiskers on a semiconductor surface specially patterned with a fine grid of silica. But the fertilizer for the forest in the latest work is gold, evaporated onto a semiconductor surface to give a partial layer only one atom deep. The authors suspect that under their experimental conditions, the gold film breaks up into fine molten beads alloyed with indium. These beads provide active sites for the arsenic- and indium-containing chemical vapour to attach to, and ride atop the growing whisker, in a mechanism first described by R. Wagner and W. Ellis 30 years ago. The density and diameter of the whiskers can be controlled through the epitaxial conditions. The hope is that the peculiar quantum properties of the whiskers will lead to useful effects, especially optical ones. The simplicity of the method makes progress here even more likely. R.P.

accessible, but human dopamine cells older than 9 weeks do not survive as well in immunosuppressed rats as younger cells do^{8,9}. Delaying the time between the abortion and the transplantation has many advantages, including the ability to test the tissue more thoroughly for safety and quality before implantation. All available experimental data clearly show, however, that transplanted dopamine cells survive and function best in the rat when the tissue is fresh. Agreement on conditions that optimize cell survival while allowing for adequate testing of the tissue should be established, and only those procedures should be used until better ones are developed.

The answer to the second question — where to put the tissue — is also generally agreed on. Dopaminergic neurons project to the basal ganglia, the site chosen for implantation. But, on closer inspection, differences lie in the details. The mesencephalic dopamine neurons project axons to the caudate and to the putamen. These two target areas are severely depleted of dopamine in Parkinson's disease (the putamen is more extensively depleted than the caudate¹⁰), and the accepted strategy has been to graft the fetal cells ectopically to these

regions, where cell survival and axonal outgrowth have been well documented in rodent and monkey studies. Spencer *et al.* implanted blocks of tissue, soon after being brought to room temperature, from one fetus into each patient unilaterally in the head of the caudate. Freed *et al.* implanted tissue from one fetus into each of four patients and from two fetuses into three patients. Some patients received unilateral implants in the caudate and putamen, while others received bilateral implants in the putamen only. Widner *et al.* implanted their two patients bilaterally in the caudate and putamen, except for patient 2, who did not receive a left caudate implant. Strikingly, Widner *et al.* used tissue from three to four fetuses in each hemisphere, that is six to eight fetuses per patient. Given the expected reduced survival of fetal tissue in the adult and aged brain relative to the developing brain, and the importance of both the putamen and the caudate in Parkinson's disease, more tissue in more areas might increase the likelihood of improved results.

Some other design characteristics of the individual studies provide important insights. In particular, Spencer *et al.* cleverly used a control group of similar

subjects assigned at random to a one-year delay in surgery, allowing for the first time a comparison of transplanted patients with a reasonable control group. Freed *et al.* used immunosuppression on every other patient and concluded that immunosuppression may be inadvisable and possibly deleterious. Finally, Widner *et al.* have followed their patients for more than two years, and the most significant results did not begin to appear until the second year, so greater success may be forthcoming as all the investigators continue to test their patients. Thus, these design improvements, in addition to the number of implanted dopamine neurons, the age of the fetal tissue, the sites of implantation, and the health of the tissue at the time of implantation, are all variables that can be changed in future studies.

Great caution should be exercised in extrapolating the greater success in the Widner *et al.* study in patients with parkinsonism induced by MPTP (1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine) to Parkinson's disease. The MPTP patients were younger and animal experiments have shown that fetal cells may survive better in younger host brains. Perhaps more importantly, however, Parkinson's is a progressive disease and may interfere with the survival of fetal dopamine cells, whereas MPTP toxicity is presumably an acute event not likely to interfere with graft survival years later. However, both these caveats provide all the more reason to optimize the chance of cell survival by putting more healthy cells into a greater number of target sites. On the basis of the results of these three studies, the long-term follow-up of these patients and other human studies now underway, improvements can and should be designed into the next set of human experiments. Although these results provide some hope that therapeutic benefits can be achieved by grafts, if further improvements over those presented here in Parkinson's disease are not forthcoming, then clinical investigations with humans should stop until further methodological advances are developed. □

Fred H. Gage is in the Department of Neurosciences, School of Medicine, University of California, San Diego, La Jolla, California 92093-0627, USA.

1. Spencer, D. D. *et al.* *New Engl. J. Med.* **327**, 1541–1548 (1992).
2. Freed, C. R. *et al.* *New Engl. J. Med.* **327**, 1549–1555 (1992).
3. Widner, H. *et al.* *New Engl. J. Med.* **327**, 1556–1563 (1992).
4. Björklund, A. & Stenevi, U. *Brain Res.* **177**, 555–560 (1979).
5. Perlow, M. J. *et al.* *Science* **204**, 643–647 (1979).
6. Lindvall, O. *et al.* *Science* **247**, 574–577 (1990).
7. Freed, C. R. *et al.* *Arch. Neurol.* **47**, 505–512 (1990).
8. Stromberg, I. *et al.* *Neurosci. Lett.* **71**, 271 (1986).
9. Brundin, P. *et al.* *Exp. Brain Res.* **65**, 235–240 (1986).
10. Kish, S. J., Shannak, K. & Hornykiewicz, O. *New Engl. J. Med.* **318**, 876–880 (1988).

Merging at the magnetopause

Alan Rodger and John Dudeney

THE magnetopause is an electrical current sheet that forms where the dynamic pressure of the solar wind is balanced by the magnetic pressure of the Earth's field. It acts as the outermost, but somewhat pervious, boundary to our environment. Under normal conditions, it lies about 10 Earth radii (64,000 kilometres) above the Earth's surface in the direction of the Sun. When the interplanetary magnetic field (IMF) has a southward component, considerable energy is transferred across this boundary into the magnetosphere by a process known as merging. A deep controversy has arisen over whether the merging occurs continuously or intermittently. Episodic behaviour is indicated by important new results from studies of transient disturbances in the polar ionosphere, reported by Lockwood *et al.*¹ on page 424 of this issue.

Typically 10^{11} watts are transferred through the magnetopause during merging events (see figure), eventually to be deposited in the high-latitude ionosphere — that part of our atmosphere between 70 and 700 kilometres altitude. This power often exceeds that due to solar radiation, so that it plays a great part in determining the structure and dynamics of the ionosphere-magnetosphere system. Understanding these outer parts of the Earth's atmosphere is crucial to building a consistent quantitative model of the magnetosphere and its responses to external forcing from the Sun. Such a model would be of considerable practical value because of the impacts that magnetospheric variability has on modern, technological society. These impacts include disruption of service from satellites (all of which orbit within the ionosphere and magnetosphere), damage to and premature loss of satellites, and disruptions to terrestrial power and communications systems.

The physics of the merging process is poorly understood. One difficulty is the lack of *in situ* satellite observations of the process. Much of the current understanding is inferred from low-altitude ionospheric signatures measured by satellites and ground-based instruments. One universally accepted signature of merging

is the satellite observation in the magnetosphere and polar ionosphere of energetic electrons and protons with very similar energy spectral characteristics to those observed in the magnetosheath. The energetic electrons take just a few seconds to travel from the dayside magnetopause to the ionosphere, whereas the more massive protons take several minutes; those with greatest field-aligned velocity arrive first. With steady merging, a simple dispersion signature would be observed with the mean and lowest ion energies decreasing with increasing latitude. But often, latitudinal structure is observed in this ion spectrum; it is the occurrence of this that Lockwood *et al.*¹ assert is the signature of intermittent merging at the magnetopause, which they attribute to a 'pulsating cusp'.

Single satellite measurements alone cannot distinguish changes due to the movement of a plasma structure from those due to its evolution. Lockwood *et al.* use ground-based measurements of elevated electron temperature, one of the many consequences of the precipitation of magnetosheath particles into the

atmosphere, to resolve spatial and temporal changes. They identify three separate events in rapid succession, which they argue is consistent with a pulsating cusp.

Previous satellite and ground-based observations indicate that the mode value of transients associated with merging is 8 minutes, but that the distribution of timescales is very wide. It takes tens of minutes for the slowest ions to reach the ionosphere, so there is a large probability that another transient will have started before the last one has completely died away. Lockwood *et al.* point out that the ionospheric effects of these transients will be abutted to each other: a series of transient events at the magnetopause will produce their observed complex spectrum of energetic protons. If transients occur with sufficient frequency, the characteristics of quasi-steady-state merging would be observed at ionospheric altitudes, namely simple monotonic ion dispersion signatures.

How secure is the pulsating cusp model? It predicts that the equatorward boundary of magnetosheath-like precipitation moves along the latitudes at ionospheric heights on timescales similar to that of merging at the magnetopause. Such motion has been detected by the sophisticated coherent and incoherent

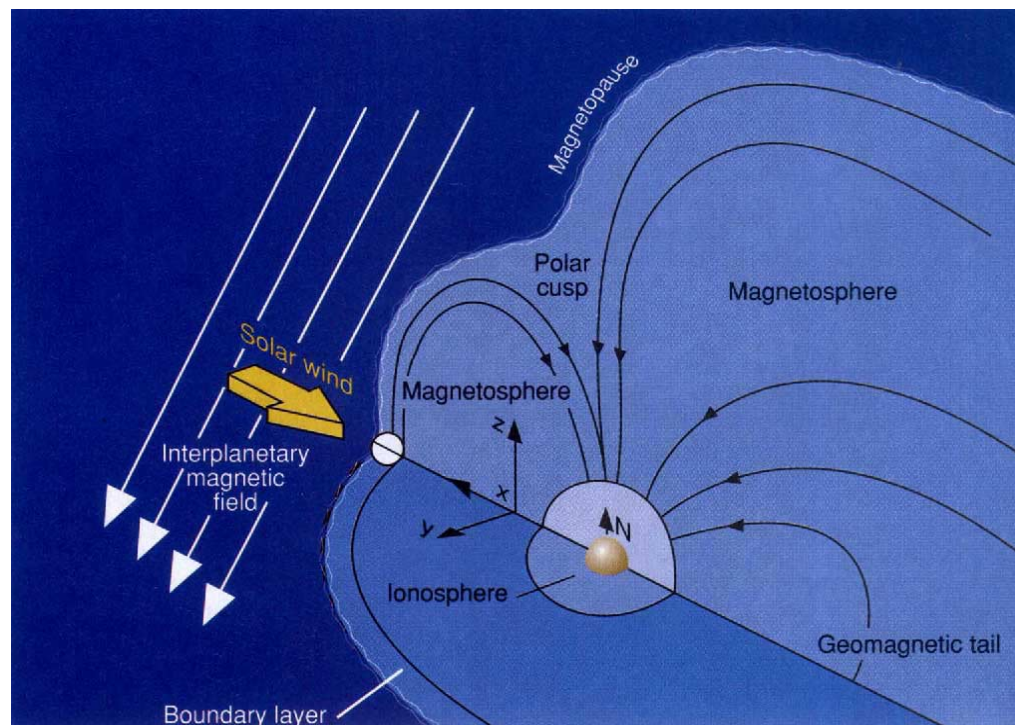


FIG. 1 Structure of geospace comprising the solar wind with the interplanetary magnetic field (IMF) embedded within it, the magnetosphere, and the ionosphere. The magnetopause lies about 64,000 kilometres towards the Sun (positive x -direction). The magnetosheath is the region immediately upstream of the magnetopause. The geomagnetic tail, only part of which is shown, extends far beyond the Moon's orbit to the right. The small circle on the magnetopause shows where merging is likely to occur when the IMF has a southward component (B_z negative). Under these conditions, considerable energy is transferred from the solar wind, through the polar cusp and into the magnetosphere. It is finally deposited in the ionosphere, the region between 70 and 700 kilometres altitude above the Earth's surface.

radar systems: observations that support the Lockwood *et al.* model. But the searching of more radar data is necessary to determine if the signature of continuous merging occurs, namely a steady equatorward edge to the precipitation boundary.

For the merging to occur, the critical factor is the orientation of the magnetic field of the magnetosheath, rather than that of the IMF although the two are related. Even with a steady IMF, dynamic and spatial variations of solar-wind pressure will result in changes in the magnetosheath and the magnetopause. These changes manifest themselves as transient behaviour in the ionosphere^{2,3}, but fortunately their ionospheric signature is sufficiently different from that of merging to be identifiable. The ionospheric flows caused directly by pressure pulses are weak relative to those caused by merging, demonstrating that they are less effective in dissipating solar wind energy into the ionosphere.

But they may play an important catalytic role for merging. Do pressure pulses cause 'roughness' in the profile of the magnetopause which in turn leads to transient merging? Would merging be continuous for steady IMF conditions, were it not for these pressure variations? Do intrinsic properties of the magnetosheath and the magnetopause affect whether merging is continuous or intermittent? These are major unresolved questions which affect both models of merging.

One hurdle to observations is plasma turbulence near the magnetopause and at the ionospheric altitudes. Under these conditions, it can no longer be assumed that the electric field near the merging site is related simply to that seen in the ionosphere.

Tension

Most merging events studied using ground-based data have been for conditions when the y-component of the IMF (see figure) was greater than 5 nT (more than 50% greater than normal). These events produce the most dramatic ionospheric effects because of the release of field line tension, as the merged field line is swept azimuthally before being dragged down the geomagnetic tail by the solar wind. Additional currents flow between the magnetosphere and the ionosphere, large plasma jets³ occur (often 100 kilometres wide and 1,000 kilometres long with velocities exceeding 2 kilometres a second) in the ionosphere, and intense red (630 nm) and green (558 nm) auroral emissions⁵ are observed. The latter are the only readily identifiable manifestation of magnetospheric dynamics to the casual observer.

Optical imaging of the aurora can be used to study the spatial and temporal

variations of cusp precipitation. One significant complication is that the height of the red and green aurorae is assumed to be 240 and 110 kilometres altitude respectively. But given the rather special circumstances occurring in the ionospheric footprint of the cusp, the altitude at which these emissions occur is probably highly variable in space and time. Such variations lead to serious errors in determining the location of the auroral features. Triangulation by observations from pairs of sites would obviate the need for any assumption, allowing the spatial relationship between the various ionospheric signatures of merging to be established beyond doubt. Thus optical studies will play a crucial role in resolving the dispute between the two models of merging.

Complex spectra

Lockwood *et al.*¹ use a relatively simple ion spectrum to support their model of the pulsating cusp. Significantly more complex spectra are often observed, some of which exhibit rising energy with increasing latitude. Further, the energetic electron spectra in the cusp region have not been studied in detail. These spectra are highly structured with bursts which are sometimes related to the discontinuities in the ion spectra but often not. Such features cannot be explained either by the pulsating-cusp model in its present form, or by continuous merging.

The pulsating-cusp model is a powerful hypothesis, but problems remain. The prospects for resolving the controversy are excellent. The NASA/ISAS (Japanese Institute of Space and Astronautical Science) Global Geospace Science Mission successfully launched last year the first of three satellites intended to investigate the main energy sources and sinks in geospace. CLUSTER, an ESA initiative to provide a small constellation of satellites ideal for *in situ* measurements of merging and the magnetosheath, will be launched in two years. New ground-based facilities, such as the Svalbard EISCAT radar, the SUPERDARN high-frequency radar network, and magnetometer networks, will all be ideally suited for studying the spatial and temporal variations of the cusp. Together with computational advances, these should pave the way for a merging of opinions on this critical region of geospace. □

Alan Rodger and John Dudeney are at the British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

1. Lockwood, M. *et al.* *Nature* **361**, 424–428 (1993).
2. Sibeck, D. G. & Newell, P. T. *Eos* **72**, 475 (1992).
3. Glassmeier, K.-H., Hoenisch, M. & Untiedt, J. *J. geophys. Res.* **94**, 2520–2528 (1989).
4. Pincock, M. *et al.* *J. geophys. Res.* (in the press).
5. Sandholt, P. E., Lockwood, M., Lybæk, B. & Farmer, A. D. *J. geophys. Res.* **95**, 21095–21109 (1990).

DAEDALUS

Save the Earth!

THE Sun is steadily getting hotter. By subtle geochemistry, this reduces atmospheric carbon dioxide. K. Caldeira and J. F. Kasting (*Nature* 360, 721; 1992) reckon that a billion years from now, when the Sun is about 10 per cent hotter than at present, plant photosynthesis will cease for lack of CO₂, bringing life on Earth to an end.

Various proposed high-tech fixes include huge mirrors in space to deflect sunlight, planet-scale decomposition of carbonate rocks, even a mass exodus to Mars, made habitable by further planet-scale engineering. No global society could last long enough to complete such gargantuan schemes. For, says Daedalus, all societies believe nonsense. It spreads irresistibly through conformity, and changes slowly through fashion. As in cell replication, any given society sooner or later hits on a chunk of lethal nonsense, and collapses. In a world of many independent societies, this does not matter. But once communications technology has formed a single global society, the time must come when everyone believes the same lethal nonsense simultaneously. Civilization then collapses globally, leaving no seed from which it could rise again. (Communism is a good example of such nonsense, lethal in a mere 80 years. Current candidates include the desire for endless economic growth.)

So once in place, a counter against solar warming must operate automatically for millions of years, despite the extinction of the civilization (or even species) that devised it. Daedalus's counter is simply to move the Earth away from the Sun. A mere 5 per cent increase in its orbital radius over the next billion years should do the trick. This could be achieved, neatly and automatically, by insulating the Moon with silica aerogel.

This material is porous, chemically inert, transparent to sunlight, but a wonderful heat insulator. It would encase the Moon in the ultimate greenhouse, warming it relentlessly by trapping solar and internal energy. Soon the more strongly heated sunward side would start to evaporate through its porous 'Moon-cosy'. The resulting vapour pressure would push the Moon away from the Sun. Bound in their common orbit, the Earth would come too. Daedalus calculates that a mere few millionths of an atmosphere of lunar vapour pressure should propel the Earth outwards fast enough to save it from solar overheating. Whenever advanced societies arose, of whatever species, they would invent space technology, discover the ancestral Moon-cosy, realize its purpose, and carry out any maintenance and updating before they collapsed again.

David Jones

Radioactivity of snake venom

SIR — A considerable part of the former Soviet Union is contaminated with various radioactive isotopes because of planned release of radioactive waste from nuclear fuel plants, accidents at atomic power stations, surface and underground nuclear explosions, and so on. Information about such contamination was kept secret, so people did not (and often still do not) know that they are living on contaminated territories.

To our knowledge, the collection of snake venom from natural populations is significantly increasing in the former Soviet Union. The only venomous snake that is widespread in the main part of the European zone, the Urals and Siberia, is the viper (*Vipera berus*). Its venom is usually collected using primitive methods and is delivered to foreign consumers through a chain of agents with no state control. But because of the environmental contamination, it is highly likely that the snake venom is contaminated with radioactivity.

There are small amounts of Co^{2+} in *V. berus* venom¹. The high Ca^{2+} content in venoms of all studied species of snakes², and the similarity of Ca^{2+} and Sr^{2+} ions in metabolic reactions, are well known. Reptiles are able efficiently to retain Sr^{2+} from the environment: the accumulation coefficient measured using ^{90}Sr is as high as 150 for reptiles (compared with 20–50 for birds, and about 10 for mammals)³ with respect to soil. Thus, ^{90}Sr seems to be a likely contaminant of snake venoms.

It is simple to measure ^{137}Cs , ^{90}Sr and ^{60}Co isotopes in venom as it is necessary only to count the radioactivity of dry venom samples. Because venom contains K^+ ions, the natural ^{40}K isotope will contribute to the total radioactivity, but differential quantification of ^{40}K , ^{60}Co , ^{90}Sr and ^{137}Cs isotopes is possible because of their different radiation spectra. Less than 10^{-11} Ci of each isotope can be detected using a routine radioactivity counting procedure. Because the Co^{2+} concentration in *V. berus* venom exceeds 3 μg per g of dry venom, the detection threshold for ^{60}Co in venom corresponds to a specific radioactivity below 10^{-4} Ci mol^{-1} . The concentration of unbound Co^{2+} in the environment is usually small, its concentration in dry venoms being about three orders of magnitude higher than in the waters of snake habitats.

The lack of data on the presence of Sr^{2+} in the venom hinders similar calculation for ^{90}Sr . However, given the high levels of ^{90}Sr contamination mentioned above, this isotope could well be present in venoms. Similarly, ^{137}Cs usually dominates in radioactive out-

bursts, and is found in almost all organisms living in contaminated areas.

Since migration of *V. berus* individuals is restricted (*V. Ragozin*, personal communication), the overall contamination of an area can in principle be estimated with a high sensitivity by measuring the radioactivity of snake venom. But financial support for ecological programmes in the former republics of the Soviet Union has decreased due to the general economic crisis, and our research to detect environmental pollution using venom analysis is no exception.

These considerations prompt us to ask people in the West receiving samples of snake venoms to measure their radioactivity and make the results available with appropriate additional information (territory, date of venom collection, and

so on). These results will be of great value for people living on contaminated territories and for ecological monitoring. The information could also be important for saving natural *V. berus* populations.

(After this manuscript had been prepared, we learnt that a large (2 kg) consignment of snake venom from Russia was impounded by customs officials because of its high radioactivity.)

Andrey A. Nedospasov
Institute of Molecular Genetics,
Russian Academy of Sciences,
Moscow 123182, Russia
Alexandr V. Cherkasov
Russian Research Centre,
"Kurchatov Institute",
Moscow 123182, Russia

1. Prokhorova, G. V., Osipova, E. A., Lebedeva, E. Yu. & Rodina, E. V. *J. analyt. Chem.* **46**, 2279–2282 (1991).
2. Stoker, K. (ed.) *Medical Use of Snake Venom Proteins* 33–56 (CRC, Boca Raton, 1990).
3. Romanov, G. N., Spirin, D. A. & Alexakhin, R. M. *Priroda* **5**, 53–58 (1990).

Fluctuating asymmetry measurement

SIR — There is much interest in the use of fluctuating asymmetry as an indicator of selective pressures on an ornamental character^{1–4}. In particular, it has been suggested that the degree of symmetry of a character, if it is paired, can indicate male quality as well as its overall length, and if this is so a negative relationship between overall length and degree of fluctuating asymmetry of a character would be seen across individuals. We would like to draw attention to two statistical problems in the measurement of fluctuating asymmetry that could produce some of the results published.

The first point concerns the use of 'relative fluctuating asymmetry'. This index, calculated as the absolute difference in the lengths of paired characters divided by the mean length of the character, is regressed against mean character length. But in the examples above, the same measurement, mean character length, occurs as a component of both the dependent and independent variables, which can lead to spurious correlations⁵. The results of regressions of relative asymmetry against mean character length are meaningless. To give one example, relative asymmetry can lead to significant negative correlations purely from measurement error. Take the hypothetical example of a paired character in which the two elements are always equal in length. A small but consistent measurement error assigned (presence or absence) at random to each of the characters can produce a significant negative relationship between relative fluctuating asymmetry and length even when there is no real difference in length. For purely statistical reasons relative asymmetry should not be used in

analysing the relationship between character length and asymmetry. (An assessment of asymmetry can be made from absolute measurements.)

Our second point concerns the use of absolute asymmetry, which we illustrate with data on spur length of the ring-neck pheasant *Phasianus colchicus*. Differences in character length may be due to growth rate, injury or damage, shortening the length of one of a pair of characters. This may bias the results of a regression of absolute asymmetry against character length depending on which measure of character length is used as the independent variable, particularly in the case of injury or damage where the difference in character lengths can be large.

We have data on left and right tarsal spur length of 233 male ring-neck pheasants from five estates in southern England where pheasants were bred in captivity and released to supplement wild stocks for hunting. Data were collected from birds shot during December and January 1990, aged one year or older. We ran regressions of the absolute difference in spur length against two different measures of character length (mean length of the paired character or the length of the longest character), controlling for age in each case. Age is used as a control variable here because absolute asymmetry covaries with age (absolute asymmetry (mm): first-year birds, $n=188$, $x=0.56 \pm 0.05$ (s.e.m.); 2+ year-old birds, $n=45$, $x=1.51 \pm 0.23$; t -test $P<0.001$).

When mean spur length was used as the independent variable we found a significant negative relationship (standardized regression coefficient -0.18 ,

$P=0.009$, $n=233$). But this relationship disappears when length of the longest spur is used as the independent variable (standardized regression coefficient 0.036, $P=0.61$, $n=233$). If one element of a pair of characters (a spur in this case, but it could be a tail feather) is shorter than the other because of differences in growth rate, injury or damage, then there will be an increased difference in character length and a lowered mean length which will bias any analysis towards giving a negative slope. The only rigorous approach is to use maximum length as the independent variable.

In conclusion, the index of relative asymmetry should never be used when looking for patterns of fluctuating asymmetry with character length. When measuring absolute asymmetry, the value for the longer of the two elements of a pair should be the independent variable. Any conclusions drawn from analyses not following this approach are flawed.

Matthew S. Sullivan

Department of Animal and Plant Sciences,
University of Sheffield,
PO Box 601,
Sheffield S10 2UQ, UK

Peter A. Robertson

Nicholas A. Aeblischer
The Game Conservancy Trust,
Fordingbridge,
Hampshire SP6 1EF, UK

1. Møller, A. P. *Anim. Behav.* **40**, 1185–1187 (1990).
2. Møller, A. P. *Nature* **357**, 238–240 (1992).
3. Møller, A. P. *Proc. R. Soc. B* **248**, 199–206 (1992).
4. Møller, A. P. & Høglund, J. *Proc. R. Soc. B* **245**, 1–5 (1991).
5. Pearson, K. *Proc. R. Soc.* **60**, 489–498 (1897).

Pain pathways and plasticity

SIR — Ramachandran *et al.*¹ reported that in two patients with amputation of an upper limb, cutaneous stimulation of the lower face elicited topographically mapped referred sensations in the phantom hand. These results are analogous to earlier physiological findings by Pons *et al.*² in adult macaque monkeys; after deafferentation, cortical regions that initially represented the forelimbs became responsive to facial stimulation. It is unlikely that either result is due to new anatomical sprouting and Ramachandran *et al.* accordingly interpreted their own findings as the selective unmasking of pre-existing thalamocortical projections. Pons³, however, has argued that such 'unmasking' is implausible as "anatomical studies have repeatedly demonstrated that areas of cortex representing the hand do not receive connections from regions of the brain representing the face."

Before it is concluded that the linkage is totally mysterious, we wish to draw attention to a release phenomenon that has not so far been mentioned in the current context. In 1920, Marinesco and Radovici⁴ reported evidence for a direct pathway between the palm and the lower face. In patients with organic cerebral disease (often involving the frontal lobe), scraping the palm in the radial direction from the thenar eminence elicits a brisk contraction of the ipsilateral mentalis muscle. The reflex is not pathognomic of any one disease; it can be seen in neurological disorders that include motor neuron disease, stroke, Parkinson's disease and various encephalopathies⁵. The original patient of Marinesco and Radovici⁴ had amyotrophic lateral sclerosis with extensive involvement of the corticobulbar tracts. Radovici⁶ was the first to consider that cortical proximity of the hand and face areas was a significant factor in the genesis of the reflex. This palmomental reflex (PMR) is also reliably seen in neonates^{7,8}.

The clinical significance of the PMR has often been disputed as its reported incidence in normal adults ranges from 0 to 100%! In part, these figures reflect differences due to age; the incidence of the reflex rises steadily over the range 20 to 90 years⁵. These discrepancies are also due to differing modes of elicitation and observation. Thus Bracha⁷ reported an incidence of 0% with mechanical stimulation, visual inspection, and the constraint that the reflex should be elicited on 10 rapidly successive trials; Caccia *et al.*⁹ found an incidence of 70% with mechanical stimulation but an electromyographic (EMG) response recorded from the mentoneal muscles; Reis¹⁰ reported a 100% incidence with electrical stimulation, sufficient to cause pain, directly over the ulnar or median nerves at the wrist and EMG responses in the relaxed mentalis muscle. The reflex disappears "with the loss of superficial pain during cuff compression of the brachial artery"¹⁰, and with "superficial procainization of the reflexogenous zone, leaving deep pain intact"¹¹.

It would thus seem reasonable to explore the relationship between the anatomo-physiological substrate of the PMR and the striking neuronal plasticity documented by Ramachandran *et al.*¹. That pain-pathways are involved both in the PMR and in modulating cortical plasticity^{12,13} suggests a mechanism involving A-delta fibre and/or C-fibre afferents. The short latency component of the PMR is unlikely to involve the precentral motor strip (area 4)¹¹; a "primary brainstem genesis" is proposed by Caccia *et al.*⁹. The late response has been conjectured to involve projections from the striatum to the thalamus with

the seventh nerve nuclei reached through thalamocortical pathways^{5,9}. The "unmasking" hypothesis proposed by Ramachandran *et al.*¹ may therefore apply at the brain-stem level implicated in the PMR.

John C. Marshall

Peter W. Halligan

Neuropsychology Unit,
University Department of Clinical
Neurology,
The Radcliffe Infirmary,
Oxford OX2 6HE, UK

1. Ramachandran, V. S., Stewart, M. & Rogers-Ramachandran, D. C. *NeuroReport* **3**, 583–586 (1992).
2. Pons, T. P. *et al. Science* **252**, 1857–1860 (1991).
3. Pons, T. P. *Science* **258**, 1160 (1992).
4. Marinesco, G. & Radovici, A. *Revue neurol.* **27**, 237–240 (1920).
5. Maertens de Noordhout, A. & Delwaide, P. J. *Archs Neurol.* **45**, 425–427 (1988).
6. Radovici, A. *Ann. de Méd.* **12**, 56–59 (1922).
7. Bracha, S. J. *Nerv. ment. Dis.* **127**, 91–94 (1958).
8. Caccia, M. R., Ossio, M. & Tornaghi, R. *Neurophysiol. Clin.* **20**, 455–462 (1990).
9. Caccia, M. R., Ossio, M. & Mangoni, A. *Electromyogr. Clin. Neurophysiol.* **31**, 151–156 (1991).
10. Reis, D. J. *Archs Neurol.* **4**, 30–42 (1961).
11. Blake, J. R. & Kunkle, E. C. *Arch. Neurol. Psychiat.* **65**, 337–345 (1951).
12. Wall, P. D. *et al. Nature* **295**, 691–693 (1982).
13. Calford, M. B. & Tweedale, R. *Proc. R. Soc. B* **243**, 269–275 (1991).

A thermostat in the tropics?

SIR — Fu *et al.*¹ conclude that there is no evidence for the cirrus thermostat mechanism proposed by us². The fundamental evidence for our hypothesis is the observed negative correlation between sea-surface warming and absorbed solar radiation in the equatorial Pacific during the 1987 El Niño. Fu *et al.* analyse satellite cloud data for 1984–87, but these data do not indicate a correlation between western Pacific cloud and variations in sea-surface temperature (SST) on climatically significant spatial and temporal scales. They also do not find a "local relationship between SST and large-scale cirrus properties". The sort of evidence sought by Fu *et al.* has now been provided by two studies^{3,4} based on longer-term satellite data.

Zhang³ uses a 15-year record of outgoing longwave radiation (OLR) at the top of the atmosphere as estimated from satellite measurements. The spatial resolution of the OLR data is 2.5° (about 250 km) in latitude and longitude. The 15-year mean OLR and SST data for the tropical west Pacific (20° N to 20° S and 90° E to 180° E) reveal that "the decrease in OLR with SST > 27° C is nearly linear at a rate of about 25 W m⁻² per °C"⁴. The rate of OLR variation with SST is within 20% of our estimate for the 1987 El Niño (see Table 1 of ref. 2). The observed relationship between long-term OLR and SST proves that clouds and/or their properties and SST

are positively correlated in the west Pacific.

Waliser *et al.*⁴ have analysed both OLR data and the frequency of "highly reflective high clouds" (HRC). The OLR and HRC data span 13 and 17 years, respectively. Waliser *et al.* find a significant correlation between HRC, OLR and SST in the western Pacific for $26 \leq \text{SST} < 30^\circ\text{C}$. HRC frequencies become negligible below 26°C . Large decreases in OLR are accompanied by much higher HRC frequencies with peak values for $\text{SST} \approx 30^\circ\text{C}$. Their results also empirically demonstrate that "in a highly convective regime the maximum equilibrium SST that can be supported is about 29.5°C ".

But the correlation between monthly mean SSTs and convective-cirrus cloudiness is reversed for SSTs greater than 29.5°C . Dynamical effects suppress convection, decrease cloudiness and dominate the thermostat effect in certain regions, where SSTs exceed the maximum equilibrium value⁴. But this exception actually supports the idea of the thermostat effect. Only a small fraction of the tropical Pacific, Atlantic and Indian oceans is warmer than 29.5°C ; about 50% of the tropical Pacific has an SST in the range $26\text{--}30^\circ\text{C}$; only 1% is in the range $30\text{--}31^\circ\text{C}$; and less than 0.01% is warmer than 31°C . These results pertain to monthly mean conditions. Thus, the strong positive correlation between SST and convective-cirrus cloudiness is valid over most of the warm ocean.

When SST data are averaged over a year, SSTs warmer than 31°C disappear and those greater than 30°C are rare. It could be that oceans warmer than the maximum equilibrium SST permitted by convection cannot be sustained for more than a year⁴.

These studies strongly suggest that highly reflective high clouds and long-wave radiation to space are correlated with SST on climatically relevant spatial and temporal scales in the western Pacific. These findings are at variance with those of Fu *et al.* These large-scale cloud radiation fields are linked with SST by deep convection. To support our hypothesis, we showed² that large reductions in tropical OLR (enhancements in long-wave cloud forcing) are always accompa-

nied by comparable reductions in shortwave radiation, consistent with earlier studies⁵.

Fu *et al.* suggest that variation in cloud area, instead of cloud thickness, is the dominant factor in the negative correlation. We invoked thick anvil cirrus because high clouds are required to explain the large reduction in OLR²⁻⁴, and cirrus occupies more than 75% of tropical cloud clusters⁶. The actual combination of these parameters cannot be inferred from satellite data alone. However, the validity of the thermostat mechanism does not depend on a specific combination of high cloud properties, but on the negative correlation between SST and absorbed solar radiation.

Fu *et al.* are also unable to find a strong link between SST and cloud-radiation indices when they average the data over the tropical western Pacific (25°N to 25°S) or the entire tropics for the 1987 El Niño. This is not surprising, as the warming is localized in the central and eastern Pacific between 10°N to 10°S . In addition, the dynamics of this coupled system introduce significant variability in the cloudiness. Convection and associated cloudiness in a particular region respond to a local warming in that region as well as to remotely forced dynamical effects. This is captured clearly in Fig. 1 where there are significant deviations around the mean superposed on the mean positive correlation between changes in SST and changes in planetary albedo. The albedo changes substantially even when the change in SST is zero, but contrary to the deductions in Fu *et al.*, this does not disprove the overall positive correlation in Fig. 1. Instead of showing shortwave cloud forcing in Fig. 1, we show change in planetary (clear and cloud-filled regions) albedo for consistency with Fu *et al.* Because changes in clear sky albedo are very small, the two quantities lead to the same conclusion.

To isolate the changes due to SST changes, in our paper² we plotted the product of change in cloud shortwave forcing and SST changes against the square of SST changes. This method isolates the local effects due to large SST increases. For example, when we expand the domain to 25°N to 25°S (instead of 10°N to 10°S as in ref. 2) the slope changes by only 10% and has a high

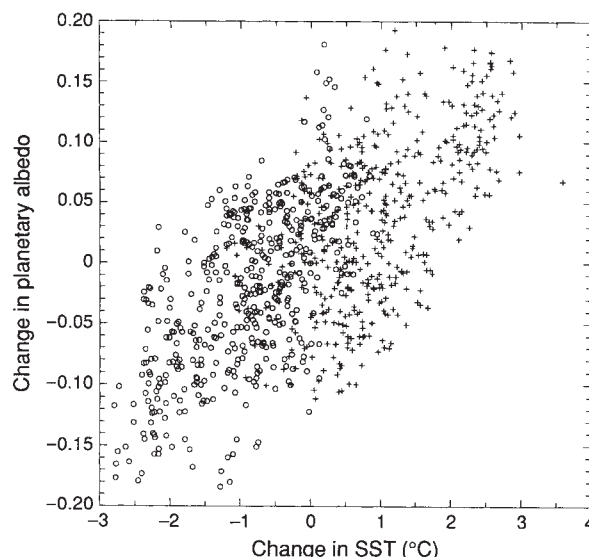


FIG. 1 Change in planetary albedo versus change in SST for the 1987 El Niño between April 1985 and 1987 (crosses) and 1987 and 1989 (circles). Each point represents a monthly mean value for a $2.5 \times 2.5^\circ$ region in the Pacific between 10°N and 10°S between 140°E and 90°W . Data sources as in ref. 2.

correlation coefficient of 0.7. Another reason for suppressing regions with small changes in SST is the uncertainty in SST data, which is of the order of 0.5°C . However, for the 10°S to 10°N domain considered in ref. 2, the usual method of correlating change in albedo against change in SST (Fig. 1) yields a slope (0.04) and a correlation coefficient (0.7) slope which are within 15% of the values obtained by our method².

The important question is: does the warm equatorial ocean respond to local changes in absorbed solar radiation (as assumed by us) or tropical mean variations (as implied by Fu *et al.*)? An answer to this question is provided by a recent three-dimensional ocean model simulation of the 1982–83 El Niño⁷. This model, which accounts for surface evaporation feedback processes, can simulate the onset and the evolution of the El Niño. However, the peak temperatures in the central Pacific exceed 32°C . The introduction of a negative cloud feedback reduces the peak model SSTs to observed maximum values of about 29°C , and the SST changes are localized in regions where the feedback is active. The radiative feedback required to limit the model SST to 29°C is applied only when SST exceeds 27°C . This process is identical to our thermostat criterion.

V. Ramanathan

W. Collins

Center for Clouds, Chemistry and Climate,
Scripps Institution of Oceanography,
University of California,
San Diego, La Jolla,
California 92093, USA

Scientific Correspondence

Scientific Correspondence is a relatively informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Because there is space to print only a small proportion of the letters received, priority is usually given according to general interest and topicality, to contributions of fewer than 500 words, and to contributions using simple language.

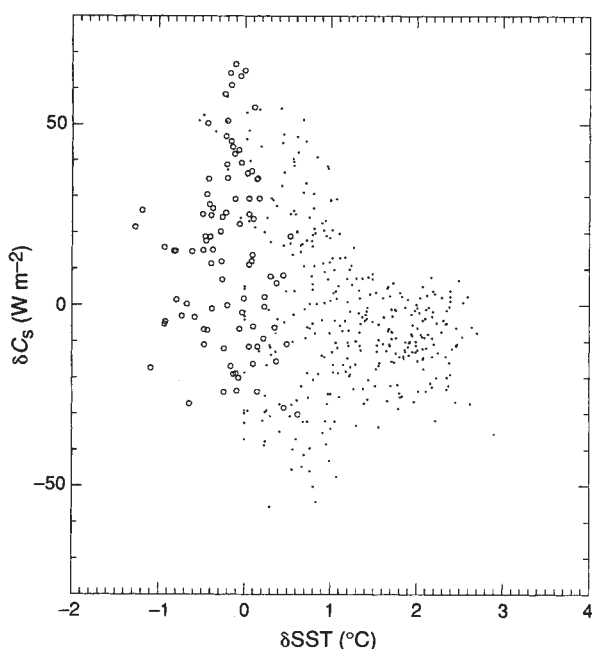


FIG. 2. Change in ERBE² shortwave cloud forcing vs. change in SST for April 1987 (El Niño) minus April 1985 in the domain 10° N–10° S. The open circles represent the west Pacific (124° E–160° E; correlation coefficient $r = -0.05$) and the dots represent the central/east Pacific (160° E–90° W; $r = -0.29$). Each point represents a difference in monthly mean values for a $2.5 \times 2.5^\circ$ region.

FU ET AL. REPLY — The existence of a spatial correlation between deep moist convection and warm sea-surface temperatures (SSTs) in certain regions of the tropical Pacific is well documented^{3,4,8}. The validity of the thermostat hypothesis² depends on whether interannual or decadal changes in SST and convective cloudiness are correlated, whether such correlations are universal, whether the correlations indicate a temporal feedback or merely reflect spatial relationships invariant with time, and whether other more important feedbacks exist. We have challenged the thermostat hypothesis on each of these counts¹. Here we present further evidence for our assertions from more recent analyses of satellite data sets.

Both Zhang³ and Waliser *et al.*⁴ argue that convection decreases with increasing SST above 29.5 °C; the latter paper also shows good agreement amongst three satellite data sets, including ISCCP. Furthermore, the variance of convection near the 32 °C upper limit of SST is fairly small. In other words, the warmest waters may not always be associated with convection at all, and must therefore be limited by some other physical process. We have shown that these warmest waters are locations of west Pacific subsidence compensating surrounding regions of intense convection on either side of the Equator⁹. Thus, the correlation between convection and SST breaks

down at the warmest SSTs, because of the non-local effect of dynamics.

Ramanathan and Collins cite correlations between ERBE shortwave cloud forcing anomalies (δC_s) and SST anomalies (δT) as the basis of their hypothesis that δC_s is a response to δT . What they actually show in their paper, however, is the correlation of the product $\delta C_s \delta T$ versus $(\delta T)^2$. We have repeated their analysis, but correlating δC_s with δT directly, using the same domain and spatial resolution of ERBE data as do Ramanathan and Collins.

We separate the data into a west Pacific segment in which small SST anomalies occur, and a central/east Pacific domain with substantial SST anomalies. Figure 2 shows that in the west Pacific, where the warmest SSTs occur, there is no correlation: $\delta T \approx 0$ while δC_s takes on va-

rious values, mostly positive but encompassing almost the full range of observed C_s variability. In the central and east Pacific, two populations exist: a $\delta T \approx 0$ cluster like that in the west, and a second group with positive δT and constant (independent of SST), slightly negative δC_s . When combined these two populations yield a negative correlation between δC_s and δT (correlation coefficient -0.29). However, this correlation is spurious because it is derived from two individual populations for which δC_s and δT are different but uncorrelated. Some of the δC_s is due to low-level cloud rather than cirrus anvil variations¹ and the magnitude of δC_s is similar in east and west Pacific. A correlation consistent with the thermostat hypothesis actually occurs over only a small fraction of the domain.

We also note that Ramanathan and Collins' Fig. 1 is the result of defining the anomaly as warm minus cold for one pair of years and cold minus warm for the other pair. Thus, points which would cluster in the same quadrant are artificially shifted into the diametrically opposite quadrant, spuriously increasing the correlation.

Furthermore, to act as the thermostat, changes in cloud should also cause a change in the upper ocean heat balance through radiative heating of the ocean. The ocean should cool significantly after solar heating is reduced by in-

creasing cloud. This would result in a positive contemporary correlation between the anomalous radiative flux into the ocean and the temperature tendency $\delta(dT/dt)$, or a positive lag correlation between the anomalous radiative flux and SST¹⁰.

However, our analysis shows that in both west and east Pacific, anomalies in temperature tendency ($\delta(dT/dt)$) are not correlated with δC_s (correlation coefficients are -0.03 and 0.02 for the west and east Pacific, respectively) in the same spatial domain as that studied by Ramanathan and Collins for the period March 1985–July 1988, consistent with Liu and Gautier's¹⁰ findings in the east Pacific for the 1982–83 El Niño. In contrast, the latent heat flux anomalies are well correlated with $\delta(dT/dt)$ over a large area of the equatorial and south Pacific (correlation coefficients 0.6 – 0.9) that covers "the area where intense warm anomalies were found during the 1982–1983 ENSO episode" according to Liu and Gautier's observations¹⁰. Anomalous increase in evaporation was found to cause a significant cooling trend and vice versa, but no significant cooling trend was found as a result of anomalous decrease in surface solar irradiance.

This evidence further suggests that changes in evaporation cannot be ignored as an important process in regulating tropical SST. We therefore stand by our original assertion that cirrus clouds are not the primary inhibitor of SST warming associated with El Niño in the tropical Pacific.

Rong Fu*

W. Timothy Liu

NASA Jet Propulsion Laboratory,
Pasadena,
California 91109, USA

* Also at

Department of Atmospheric
Sciences,

University of California,
Los Angeles, California 90024, USA

Anthony D. Del Genio

William B. Rossow

NASA Goddard Institute
for Space Studies,
2880 Broadway,
New York,
New York 10025, USA

1. Fu, R., Del Genio, A. D., Rossow, W. B. & Liu, W. T. *Nature* **358**, 394–397 (1992).
2. Ramanathan, V. & Collins, W. C. *Nature* **351**, 27–32 (1991).
3. Zhang, C. *J. Climate* (in the press).
4. Waliser, D. E., Graham, N. E. & Gautier, C. *J. Climate* **6**, 331–353 (1993).
5. Chertock, B., Frouin, R. & Somerville, R. C. *J. Climate* **4**, 639–650 (1991).
6. Houze, J. D. *J. Meteor. Soc. Japan* **60**, 396–409 (1982).
7. Harrison, D. E. *J. Climate* **4**, 539–549 (1991).
8. Fu, R., Del Genio, A. D. & Rossow, W. B. *J. Climate* **3**, 1129–1152 (1990).
9. Fu, R., Del Genio, A. D. & Rossow, W. B. *J. Climate* (submitted).
10. Liu, W. T. & Gautier, C. *J. geophys. Res.* **95**, 13209–13217 (1990).

Mr Science, warts and all

Roy Porter

The Invisible Man: The Life and Liberties of H. G. Wells. By Michael Coren. Bloomsbury: 1993. Pp. 240. £20. (To be published in the United States by Atheneum later this year.)

ONE of this century's great teachers and novelists, H. G. Wells was blessed with superabundant energies — was, indeed, a driven person. As the classic self-made man who worshipped his maker, he possessed omnivorous appetites and pursued astonishingly broad interests. In women and affairs of the mind alike, he spread himself widely if not always wisely. Incurribly ambitious, he made himself the consummate self-publicist in an age of mass communication.

Born in 1866 into a shabby shopkeeping family from Bromley and backed by a powerful mother, the young Wells developed a passionate love of writing and debating that lasted all his life. Trapped early on as a draper's apprentice and school-teacher — episodes chronicled in autobiographical fictions such as *The History of Mr Polly* (1910) which were destined to be his lasting testament — he fled suburban tedium to enjoy what proved to be his most formative experience. By studying science at Imperial College, London, sitting at the feet of the divine T. H. Huxley, he acquired the obsession with science that fuelled his whole career. His goal was to emulate Darwin's Bulldog, to be the sage and seer for the modern world. Surpassing his mentor, Wells displayed extraordinary talents for journalism and fiction. Above all, in works such as *The Time Machine* (1895), he virtually invented a new genre of popular novel, grafting scientific and technological futurism onto political utopianism in tales that voiced the hopes and fears of common men.

Flirting with Fabianism, Wells-the-thinker put his finger on the pulse of the Edwardian age. Victorian values were *passé*. The new century wanted populism and radicalism; it fostered the cult of everyman, it craved the sexual experimentation that Wells personally indulged in and discussed in his essays. Not least, the Edwardian mind looked to science to create a new social order — rational, productive, efficient, hygienic and healthy — while enjoying the frisson of fear sparked by boldly going (at least in the mind) where no man had gone before.

Wells met these needs with a blizzard of novels and tracts. Even more than Shaw, he became the mouthpiece of progressive thinking, prophesying better futures. And then, as uncertainties and forebodings grew after the Great War, Wells's own message and mood dark-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Armchair visionary — Max Beerbohm's caricature of Wells.

ened. Early faith in a new samurai expert class had evaporated by the 1930s, when the shape of things to come seemed the dynasties of the dictators. Wells sank into dejection. His last work, published in 1945 just months before his death, was aptly called *Mind at the End of Its Tether*. Wells, remarked his arch-critic, Malcolm Muggeridge, had been a naïve crusader for shallow scientific rationalism; Hitler and the Holocaust were his chickens coming home to roost.

Wells's personal life also involved a fall from pleasure to pain. Love affairs, notably with Rebecca West, went sour and created conflict. A man ever craving sexual adventure and female admiration, he found ageing hard to endure, growing testy, impatient and moody. But for all his faults, he never entirely lost the more

amiable side of his character — a gusto for life, a puckish charm, an openness to experience, a certain honesty.

In this breezy yet informative biography, Michael Coren gets the blend about right. As his choice of title indicates, he believes the true Wells has been neglected. Until now, we have had hagiography not biography. He aims to put the record straight and to give us for the first time the whole Wells, warts and all. Coren's hero is thus deeply self-centred, and perhaps a frankly selfish user of women. Yet Coren is not as original in this as he likes to think: readers of Ruth Brandon's witty *The New Women and the Old Men* (Secker and Warburg, 1990) will already be familiar with the blindspots of Wells's special style of one-way permissiveness.

The real shortcoming of this biography is that in stressing Wells-the-man Coren neglects Wells-the-mind. Alongside his hero's love of sex we could have done with more of Wells's love of science. Coren never analyses Wells's scientific beliefs in any depth: just what combination of social evolutionism and eugenism was Wells advocating? How far, indeed, was Wells actually advocating a particular scientific future or merely forecasting its imminence? Coren is inclined to paint Wells as a bit of a scientific buffoon, repeatedly outsmarted in evolutionary controversies, even by Hilaire Belloc. But this is to miss Wells's remarkable magnetism as a scientific preacher and teacher. No one as scientifically inept as Coren's Wells would have succeeded in persuading Julian Huxley to collaborate with him in 1933 on *The Science of Life* (1930), a massive textbook of popular biology that complements Wells's earlier best-selling *A Short History of the World*.

Coren fails to convey Wells's importance as a public figure. Anti-Christian and ferociously anti-Catholic, Wells viewed science as the new religion. Identification with scientific salvation made him the great prophet of the age, the thrilling, favoured author for millions of clerks and commuters. Wells's role as Mr Science deserves closer study. Ironically, his own extraordinary life, *An Experiment in Autobiography*, remains the most penetrating and poignant introduction to the man and his mind. □

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

One world

Timothy O'Riordan

The Environment in International Relations. By Caroline Thomas. *Royal Institute for International Affairs, London: 1992. Pp. 291. £25, \$42.95 (hbk); £12.50, \$21.95 (pbk).*

Global Environmental Change and International Relations. Edited by Ian Rowlands and Malory Greene. *Macmillan: 1991. Pp. 202. £42.50 (hbk); £15.99 (pbk).*

The International Politics of the Environment. Edited by Andrew Hurrell and Benedict Kingsbury. *Clarendon (Oxford University Press): 1991. Pp. 492. £40, \$72 (hbk); £15.95, \$19.95 (pbk).*

Conversion and the Environment. Edited by Neils P. Gleditsch. *International Peace Research Institute, Oslo: 1992. Pp. 337. NK240, \$40.*

Verification Report 1992: Yearbook on Arms Control and Environmental Agreements. Edited by J. B. Poole and R. Guthrie. *Verification Technology Information Centre, London: 1992. Pp. 372. £25.*

Environmental Legacy of the Former Soviet Republics. By Ruben A. M. Natsakanian. *Centre for Human Ecology, University of Edinburgh: 1992. Pp. 209. £20 (pbk).*

Paradise Deferred: Environmental Policymaking in Central and Eastern Europe. By Duncan Fisher. *Royal Institute for International Affairs, London: 1992. Pp. 82. £10, \$20 (pbk).*

THERE is no doubt that environmental issues are as much if not more an international matter than a domestic one. For example, although toxic waste dumping is a purely national problem in the sense that the waste disposal and the supporting policy are indigenous to one nation, the ramifications do cross borders. Cleanups involve international, or at least bilateral, aid if a country is poor. And in any case, the resulting public ill-health is not just a moral issue: it may well cause desperate people to migrate across national boundaries in search of environmental and economic salvation.

These are serious matters that no nation state can afford to ignore. But obviously of greater relevance to international affairs is the response to two particular classes of environmental dilemma. One is the management of resources that are part of the common heritage of humankind, such as migratory species (whales, wading birds) or unique species or habitats (the white rhinoceros, the moist tropical forest, the outer shield of ozone that protects us from excessive ultraviolet light). The other is the control of emissions, which

may appear in one part of the world but create universal damage when they accumulate in the fluxes of chemicals and energy that bathe Earth's life.

International action to safeguard pollution of the high seas, to stop the trade in endangered species, to limit the hunting of whales and to control the exploitation of seabed resources has been provided for in a host of treaties, conventions and protocols over the past 20 years or so. Much of what has to be tackled in international common management has legal and political precedents. The first three of these books provide a great service in reminding readers that the complexities of regime formation and treaty provision are grounded in rich experience of arms negotiation, security protection and resource allocation.

What is evident from this history is the slow pace of reaching agreement, the unwillingness to trade national interest for some collective good unless the national interest is clearly safeguarded, and the deep suspicion of new agencies and regulatory entities that limit room for manoeuvre and create self-serving bureaucracies. The essays in the volume edited by Hurrell and Kingsbury are here of great value, for they provide an excellent perspective on how international action is developed, and clarify the important distinction between the hard law of binding treaties and legal obligations, and the so-called 'soft law' of custom, consensus, flexible arrangements and nonbinding institutions of short shelf-life but potentially long-term commitment.

The first three volumes each provide a valuable review of the difficulties and opportunities of preparing appropriate regimes for the management of the 'big' environmental issues such as ozone depletion, climate change, habitat loss and resource management across national borders. This is where soft law is most effective, aiming to nudge suspicious states into voluntary schemes and pilot projects that seem to involve no commitment, but which create international expectation of political consensus and momentum. Thus the Montreal Protocol on the removal of gases that deplete the outer ozone layer has provisions for a regular review through which the gases selected and the timetable for their removal can be brought up to date. The role of nongovernmental organizations in this process is vital, because they too act internationally, are scientifically well connected, and serve as a powerful educative force in this day and age of environmental journalism. Both Hurrell and Kingsbury's and Rowlands and Greene's volumes cover this aspect well for ozone loss and biodiversity.

The main problem with soft law lies in

the interaction among scientific discovery, scientific proof, the scope for establishing a common purpose and the appropriate mechanisms for ensuring that states act according to their culpability. This is an especially difficult problem for the removal of the so-called greenhouse gases, because states still lagging in the process of economic advance claim that they should be left alone while the long-term polluters (that is, the rich industrial nations and the impoverished former Soviet bloc) get on with removing their carbon dioxide.

Sadly, none of the three books takes a close look at the subtleties of the interactions among collective action, scientific evidence, economic development and employment. This is precisely what is plaguing climate negotiations at present. And it is also a key factor in the determination of the shape of the United Nations Commission on Sustainable Development, formed at the UN Conference on Environment and Development in Rio de Janeiro in June last year but still embryonic.

Hurrell and Kingsbury show vividly how much suspicion there is about any new UN agency. They also emphasize the dangers of trying to push soft law into hard law too quickly, no matter how urgent the environmental dilemma. They correctly point out that for most negotiators the environmental movement is still fairly small beer. Mechanisms for transferring the funds of rich countries to poorer countries to offset global danger, namely the Global Environment Facility of the World Bank and the UN Development and Environment Programmes, are still very embryonic; they have only limited scope for the vast job of transferring resources and technology to the South that is ultimately required. 'Softly softly' procedures may do it, but the pace will be agonizingly slow. This is because so much linkage to other pertinent issues must also be negotiated and in a way that seems at the outset to be nonthreatening.

Conversion and the Environment and *Verification Report 1992* provide some sober reading on the environmental problems associated with the disposal of military assets and on the difficulty of verification of international arms agreements, even when a lot is at stake if such verification is not undertaken. Who knows how far verification and enforcement will be possible in sustainable development plans.

Similarly, and of equal foreboding for ineffective international action, are the two reports on the environmental calamities of Eastern Europe. One should not be too dismayed, for at least much is now known that was almost wholly suppressed before. And some efforts are being made to fund support for non-

governmental organizations in self-help schemes and in drafting appropriate legislation. Similarly, scientific exchange and the sharing of equipment and monitoring schemes should begin to harmonize environmental reporting over the next decade. For all this, one must be impressed at the rapidity of movement.

But the scale of environmental devastation in Eastern Europe is enormous, with as much as one-fifth of inhabited areas seriously polluted to the point of cutting short the life expectancy of up to half the children. The possibility of full recovery is slim, at least until collectively we recognize that in environmental renaissance there is the scope for economic recovery that could put whole nations back to work. Peace and environmental transformation go hand in hand. It is always wise to remember that global environmental management is not a stand-alone issue. All these books, in their fascinating ways, reveal this essential but sobering truth. □

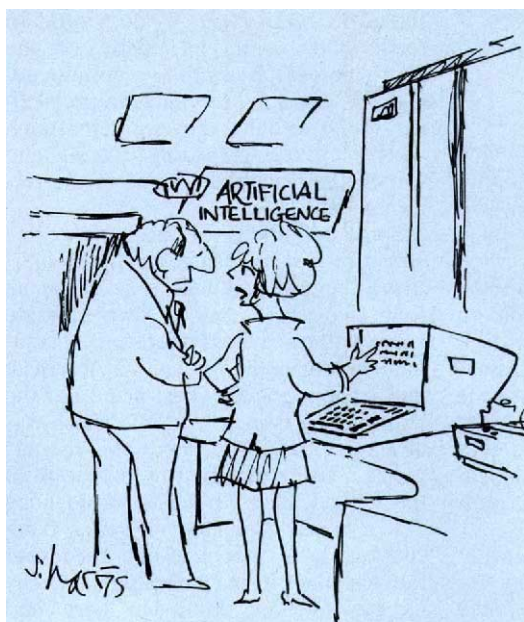
Timothy O'Riordan is in the Centre for Social and Economic Research into the Global Environment, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK.

Of two minds

H. J. Eysenck

Wolfgang Pauli und C. G. Jung: Ein Briefwechsel, 1932–1958. Edited by C. A. Meier. Springer. Pp. 275. DM98.

THIS collection of letters in German is a curious book, just as the association between an outstanding scientist such as Pauli and a mystic such as Jung is rather curious. The association came about because of Pauli's apparent need for treatment, although it is never very clear what exactly he was suffering from. Pauli sometimes regards Jung as a leader in the field of psychology, and a natural scientist (*Naturwissenschaftler*), rather than a *Geisteswissenschaftler*, but it is difficult to understand how he could make such an error. Freud at least tried to abide by the rules of science, although not very successfully; Jung never bothered, and retained the mantle of 'guru' and mystic right to the end. All his work is impressionistic and anecdotal; there is no attempt at rigorous statement of hypotheses, deduction of testable consequences and empirical proof or disproof. Jung simply stated his views about such things as the "collective unconscious", regarding them as facts simply on the grounds of *ipse dixit*. He was miles removed from orthodox psychology, and has only a small but devoted group of followers,



SCIENTISTS and science come under fire once again in Sidney Harris's latest collection of cartoons. *From Personal Ads to Cloning Labs* is published in paperback by W. H. Freeman, £8.95, \$10.95.

few if any trained in psychology.

Pauli noted some interesting things that might have made him suspicious. Thus he reports in one of his letters that as the analysis proceeded, more and more Jungian symbols appeared in his dreams; he failed to realize that he was just being trained to dream Jungian dreams! He also indulged in dream interpretation; when he dreamed of a wasp with differently coloured stripes he told Jung that it typified the dual aspect of his personality. So much for science. In turn Jung told him that UFOs were the "projected", that is, concretized symbolism of the process of individuation. Pauli never asked if any meaning was attached to such a statement or whether there could be any form of proof.

Pauli certainly followed Jung's teaching, and attempted to support some of his ideas by reference to physical theories. Thus he argued that the principle of complementarity presented a fundamental analogy to the notion of 'consciousness' and 'unconsciousness'. He also considered that for a unified understanding of nature it was necessary to go back to the archaic backgrounds of natural science; he mentions Kepler, to whom the Copernican system appeared as a symbol of the Holy Trinity.

Quite generally, there is a curious attempt to find analogies between psychological concepts and physical laws. Jung followed suit, but seems to have exaggerated the analogy dreadfully. Thus he misstates the Heisenberg principle ("that which is observed is a disturbance produced by the observation") and concludes that "if one looks long enough into a dark hole, one perceives what it is

that does the looking", which he then identifies with the fundamental principle of Yoga. All of this leads to a justification of clairvoyance and precognition! Thus a very remote analogy from physics is used to justify parapsychological phenomena, whose existence are taken for granted. The whole letter should be read to give an insight into the cognitive world of a mystic (pages 11–12). How Pauli came to argue that Jung was a leader in natural science and psychology passes understanding.

Most readers will find it difficult to understand much of what Jung says, or even to imagine ways of testing it experimentally. Pauli's approbation reminds me of Newton's excursions into alchemy, or his disquisition about '666' and the nature of the beast in the Bible, a curious aberration that even psychologists find difficult to explain. But then even supremely intelligent people commit imbecilities when removed from their proper field — the Webbs visited the Soviet Union and said: "We have seen the future, and it works!"; Bertrand Russell called Kennedy "more wicked than Hitler" when he stood up to Khrushchev, and later advocated a nuclear strike by the West against Russia. This particular aberration may serve as a clinical document, but it will not change Pauli's great scientific achievement, nor will it serve to shore up Jung's vanishing reputation. Its main effect will be to demonstrate to readers the vagaries of a great mind. □

H. J. Eysenck is at the Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK.

Big blunder

Brian J. Ford

Academic Press Dictionary of Science and Technology. Edited by Christopher Morris. *Academic*: 1992. Pp. 2,432. £68, \$115.

THIS is a disappointment. It claims to be the definitive, world's-biggest book on science, and with 124,000 entries, together with 124 brief "windows" essays by luminaries in their fields, it seems the perfect project for the age. But the result is uneven. Although the definitions are largely as good as one might hope, the inconsistencies and poor illustrations spoil the product.

US and UK spellings are treated without consistency. There is an entry for *Amoeba* (illustrated by a drawing) and a second, separate entry for *Ameba* (with a photograph), each with its own definition. But the (US) *Ameba* is still defined as a member of the (UK) Amoebida. *Aluminium* is listed as "the British term for aluminum", yet there are other definitions in which the convention is reversed. The plant pigment *leghaemoglobin* has an entry of its own, with the US spelling *-hem-* as a note. But from the 16 terms starting with *haem-*, haemoglobin is absent. The compound occurs only under the US spelling *hemoglobin*, whereas the UK version here warrants not even a mention.

The drawings look as though an editor had decided to lighten the load of words by commissioning a job-lot of pictures from an artist with a reputation for speed. Many show species as exemplars of a family, labelled with the single family name, which is like captioning a picture of a football boot as "Sport". There is a single swallow-tail butterfly captioned *Papilionidae*, and a perfunctory sketch of a bird louse labelled with its family name, *Laemobothridae*. A few radiating leaves are ambitiously endowed with the caption *rubber tree*, while the entry for *bird of paradise* is confusingly decorated with an otherwise unidentified sketch of a flowing *Strelitzia reginae*.

No eclipse can have a global penumbra like the one shown on p. 1,602. The doodle of a kidney is dreadful, and the *gluteus maximus* is one of the most hasty drawings I have seen in any book. The sketch that decorates the otherwise unidentified *loon* compares unfavourably with the image of a great northern diver published by Ole Worm in 1665. Robert Hooke published his study of the louse in that year, more vivid by far than the image in this book. Here, *louse* is confusingly illustrated with a dipteran fly. The subtitle of the drawing in paren-

theses — *book louse* — does little to rectify the confusion. Many of the photomicrograph captions are misleading. For example, the micrograph labelled *starch* actually shows potato starch under crossed polarizing filters, and without both facts the point of the picture is lost.

Some definitions are incomplete. The entry for *reactor* includes only the electrical, chemical and nuclear usages of the term: *bio-reactor* has its own separate entry, without a cross reference. In common with many books with an American link, Leeuwenhoek has acquired the wrong given name, "Anton", which was never used by him or any of his friends. It has become dignified by repetition in the United States, but one would hope that a newly compiled reference book could go back to sources and not repeat the mistake. It may also explain why Gregor Mendel, Moravian born and bred, is "Austrian" in this text. The

Periodic Table in the appendix omits the names that have been offered for the heaviest transuranic elements. Diligent readers will find the names perversely present only in the table of elements in order of atomic weight. At least, most of them are there; element 108, which is unniloctium, is curiously missing from both lists.

Books on science are liable to contain errors of fact or interpretation. The conventional excuses are familiar enough: in a fast-changing field, it is impossible to keep pace with progress, and so on. None of these explains the lamentable deficiencies in this book. Science deserves better, and Academic Press should look to its laurels (or "Angiospermae", as they would doubtless have captioned the picture). □

Brian J. Ford is at Rothay House, Mayfield Road, Eastrea, Cambridgeshire PE7 2AY, UK.

Building on firm foundations

J. P. Platt

Structural Geology. By Robert J. Twiss and Eldridge M. Moores. *W. H. Freeman*: 1992. Pp. 532. £22.95, \$59.95.

IN its broadest definition, structural geology encompasses everything concerned with rock deformation, from experimental and theoretical rock mechanics to large-scale tectonics, and overlaps with fields as diverse as basin analysis, seismology and metamorphic petrology. To many of its practitioners, however, it is much more narrowly defined, concerned principally with the geometrical analysis of relatively small-scale structures and with the measurement of ductile strain. Several recent texts have attempted to correct this restricted view by emphasizing the broad scope of the subject and its links with other areas. This is the best attempt that I have seen so far, in that it combines breadth with theoretical and technical rigour.

The book is divided into five sections, each of which starts with essentially descriptive chapters followed by theory. Section four, for example, consists of five chapters describing various types of fractures and faults, followed by three on stress theory and fracture mechanics. This approach leads to some repetition of material and occasionally to oversimplifications of theory in the descriptive sections; but it has obvious advantages, because, by selecting appropriate chapters, one can use the book for several courses from introductory to advanced undergraduate or postgraduate level.

The text is for the most part clearly

written and well organized, although the use of 'boxes' is excessive: some of them occupy three full pages and completely interrupt the flow of the main text. The authors pay particular attention to the meaning of words, and to precise and rigorous definitions (I don't agree with all of them, naturally!). I like the lucid discussion of the different types of model, and of scaling problems in physical modelling. The treatment of deformation theory is certainly sufficiently thorough for undergraduate courses, although the order of presentation is not always very logical (again, hindered by overuse of boxes), and the algebra is littered with unnecessary superscripts, subscripts and diacritical marks (such as circumflexes over all principal stress and strain symbols). Figures and diagrams are mostly carefully thought out and of high quality, including photographs reproduced from other publications.

The book is least successful where it steps beyond the traditional bounds of structural geology. For example, the section on geophysical techniques in Chapter 1, although well written, seems not to be made use of in the rest of the book; the section on critically tapered thrust wedges will, I think, be easily misunderstood by students; and the final chapters on tectonics are rather patchy. But overall it is a rigorous and comprehensive text, and a strong contender as a course book at a wide variety of levels. □

J. P. Platt is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

The centre of the Milky Way

Leo Blitz, James Binney, K. Y. Lo, John Bally & Paul T. P. Ho

Direct photometric evidence has established that the central bulge of our Galaxy (the Milky Way) is actually a stellar bar. The dynamics of the bar will drive material into the Galaxy's centre but recent X-ray observations suggest that much of the inflowing gas is driven out again in a high-pressure galactic wind. A small portion of the remainder appears to be falling onto an extraordinarily dense cluster of stars at the Galaxy's nucleus, which may contain a massive black hole.

FROM the birth of radio astronomy in 1932, when Jansky discovered cosmic radio waves and established that they come from the Galactic Centre¹, it has been clear that something very remarkable is taking place at the core of the Milky Way. At the long wavelengths he observed, the Galactic Centre turned out to be the brightest object in the sky, brighter even than the Sun. Subsequently, the Galactic Centre has been shown to be astonishingly bright at many wavelengths, although it cannot be seen optically because it is obscured by intervening dust. Observations at wavelengths to which the Milky Way is transparent have shown that the central regions are unlike other parts of the Galaxy, and that bizarre features seem to be the rule rather than the exception. Some astronomers have suggested that the energetic activity results from matter accreting onto a black hole at the very centre² with a mass as much as several million times that of the Sun³. Others suggest that it is due to an extraordinary recent burst of star formation⁴. Confirmation of either hypothesis would conceptually link the Galactic Centre with the centres of active galaxies, which release prodigious amounts of energy from their nuclear regions, and which seem to require either very massive black holes or huge bursts of star formation to account for what is observed. The centre of the Milky Way appears in many respects to be a quieter version of an active galactic nucleus.

What then is the nature of the extraordinary mass and luminosity concentration at the Galactic Centre? How does matter get to the centre to form whatever is there? Why is there such a plethora of exotic phenomena at the centre, and how can they be explained? Here we can only paint answers with a broad brush, but a coherent picture is beginning to emerge. The current observational evidence suggests that the luminosity of the Galactic Centre is generated by a cluster of luminous stars rather than by a black hole. The best evidence for the existence of a black hole is the presence within the cluster of a unique radio source, Sgr A*, but it is not compelling. The gas that feeds the star formation at the centre is brought inwards by the dynamical effects of a newly identified massive stellar bar. Most of the gas brought to the Centre is apparently later expelled in a galactic wind, but some with low angular momentum finds its way to the nucleus, where the gas is at a pressure at least three orders of magnitude higher than in the disk. There is good evidence that the infall occurs right down into the central parsec, but at a much smaller rate than the overall mass inflow rate, suggesting that any accretion onto a central black hole is episodic.

Our discussion proceeds from large to smaller scales, ultimately reaching the very nucleus of the Galaxy.

The Milky Way: a barred spiral galaxy

Figure 1 shows the most striking image of the Milky Way, which was produced by the DIRBE experiment on the COBE satellite⁵. At the centre one can clearly see the bulge of the Milky Way extending above and below the flat disk of the Milky Way, which is reddened by galactic dust in the same way that sunsets are reddened by dust in the atmosphere. The bulge, by contrast, contains little dust. Until recently it was assumed to be a flattened axisymmetric spheroid (that is, one which is round in the plane



FIG. 1 The central 4 kpc of the Milky Way as seen in the near infrared by the DIRBE experiment aboard the COBE satellite. (The entire picture runs from $l = 96^\circ$ to $l = -96^\circ$.) The picture is a false-colour representation of the near-infrared emission at wavelengths of 1.2, 2.2 and 3.4 μm . The reddish colour of the disk is caused by dust and the whitish region in the centre is the bulge.

of the Galaxy). There is now evidence that the bulge is elongated in a direction that lies in the plane; the Milky Way is a barred spiral galaxy.

Roughly half of all disk galaxies contain a bar, an elongated distribution of stars whose longest axis lies in the plane of the disk (Fig. 2). Barred galaxies have been recognized as a separate class of galaxies since the 1920s⁶, and in 1964 de Vaucouleurs⁷ suggested that the Milky Way contains a bar to explain the observed noncircular motions of the atomic gas at the centre. Subsequent detailed analyses of the motions of the atomic hydrogen gas in the plane supported this proposal⁸⁻¹⁰. Within the past year, new photometric and dynamical evidence has become available that clearly places the Milky Way in the class of barred rather than normal galaxies.

The photometric evidence results from an analysis¹¹ of the near-infrared emission obtained from a Japanese balloon-borne telescope launched nearly a decade ago¹². In the infrared, extinction by dust is an order of magnitude smaller than in the visible portion of the spectrum. By analysing the emission above and below the plane (the whitish regions in Fig. 1), it is possible to make a quantitative assessment of the shape of the bulge. If the bulge is axisymmetric, lines of sight that make the same angle on either side of the centre will appear equally bright. On the

other hand, if the bulge is elongated and its long axis is at some angle oblique to the line of sight (see Fig. 3), perspective effects cause equivalent lines of sight on either side of the centre to have systematically different surface brightnesses. For most lines of sight, the near side of the bar will appear to be brighter than the far side because lines of sight on the near side traverse paths with a higher volume density of stars. The near side will also appear to be thicker than the far side. Close to the centre however, lines of sight on the far side are actually brighter because the stellar densities sampled are almost the same, but the lines of sight on the far side traverse a longer path of stars. This counterintuitive result requires the differential surface brightness of the two halves of an obliquely oriented bar to show a sign reversal near the centre, and the photometrically determined near side to have a larger angular thickness. An axisymmetric bulge, as would be found in a normal (non-barred) galaxy, would show no such effects.

Both the Japanese data and newer data from the COBE satellite show all of the expected features^{11,13}. The orientation of the bar, which is currently poorly constrained by the photometry, is in the sense shown in Fig. 3. Analyses of the distributions of individual stars in the bulge have all shown¹⁴⁻¹⁶ an asymmetry consistent with the bar found in the infrared continuum and all are in the same sense as indicated by the analysis of the balloon photometry; the figure of the bar emerges clearly through the dusty murk of the galactic plane.

Kinematics of the gas in the bulge

A recent attempt¹⁷ to find a dynamical explanation for the observed motions of atomic and molecular gas in the central regions has provided independent evidence of a stellar bar forcing noncircular motions in the central regions. Let us first describe the gas and its motions. Atomic and molecular gas comprises more than 90% of the mass of the interstellar material, and are detected primarily through their emission at radio wavelengths. Atomic hydrogen (H I) is observed directly through its 21-cm hyperfine line, whereas H₂ is not directly observable and is traced primarily through the 2.6-mm rotational transition of the CO molecule. Measurements of the H I and CO lines enable one to infer for each point on the sky the column density of H I and H₂ moving with a given velocity, but the inferred H₂ column densities are cruder because of the extra interpretive step required. The interpretation of the measurements is further complicated by the indeterminacy of the distance to the element of gas that is responsible for the emission.

Because the neutral gas is strongly confined to the plane, in

a first analysis it is convenient to concentrate on a thin slice around $b = 0^\circ$, where b is galactic latitude; this two-dimensional slice through the data cube contains all available information about gas in the plane. It can be displayed in plots such as those of Fig. 4 of line-of-sight velocity against galactic longitude l .

In broad outline, these plots can be understood if most of the gas moves about the centre on circular orbits. Within $\sim 20^\circ$ of the centre (~ 3 kpc), plots for both H I and CO similar to Fig. 4 show a decline in the maximum velocities with decreasing longitude; this would arise naturally if gas were largely excluded from the central 3 kpc. However, the peak H I and CO velocities start to rise about 12° and 2° from the centre, respectively. The natural explanation of these upturns is that disks of H I and CO with radii of 1,500 pc and 300 pc, respectively, sit within the central 3 kpc hole. We refer to everything in the plane exterior to 1.5 kpc as the 'disk', and consider all material inside 1.5 kpc, whether or not it lies in the plane, to be part of the bulge.

Also shown in Fig. 4 is a longitude-velocity plot of the emission from the $J = 2-1$ transition of the CS molecule. The CO molecule typically traces H₂ at densities of $\sim 10^2-10^3$ cm⁻³, whereas CS traces the H₂ at densities of at least 10^4 cm⁻³. The detailed differences between the CO and CS maps imply that the pressure of interstellar gas in the bulge is much higher than it is in the disk. This point is discussed in more detail below.

In Fig. 4, there is emission by both H I and CO at 'forbidden velocities', that is, at velocities which do not occur in a model in which all material circulates in the same sense on circular orbits. Specifically, there is material at positive longitudes and negative velocities, and at negative longitudes and positive velocities. Such velocities would occur if gas moved on elliptical streamlines^{8,9,18-20}.

Dynamics of the bar

The key to understanding the distributions of Fig. 4 turns out to be closed ballistic orbits, which provide the framework for comprehending the general orbital structure of any potential; most orbits can be interpreted as epicyclic excursions ('librations') around an underlying stable closed orbit. At any energy there will be at least one closed orbit, and this orbit is said to parent the family of non-closed orbits that librate around it.



FIG. 2 From outside, the Milky Way probably looks rather like this typical barred spiral galaxy, NGC 4535. The bar is the broad feature across the center that connects the two bright spiral arms.

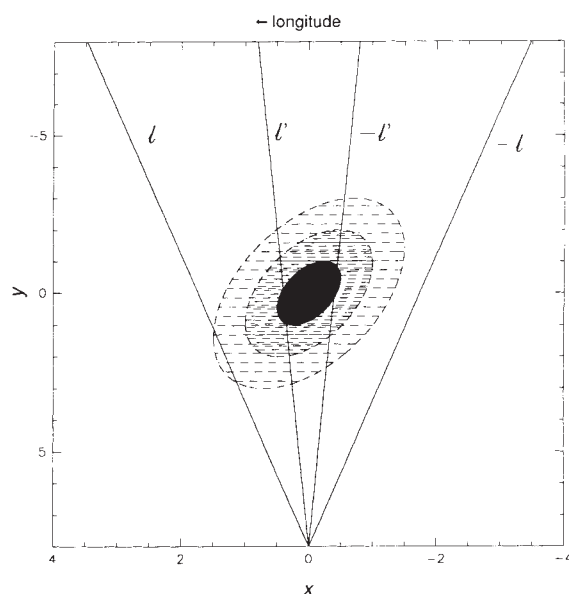


FIG. 3 A schematic bar viewed from the galactic pole showing the orientation implied by the infrared data of Matsumoto *et al.*¹². The Sun is at (0, 8). Lines of sight that make equal angles with the direction to the Centre intersect different volumes of the bar.

FIG. 5 a, Closed orbits in a barred model of the galactic potential¹⁷. The long axis of the bar is vertical and the dotted lines indicate our line of sight to the Galactic Centre. The nearly circular orbits on the outside of the left panel are outside the OLR. The orbits with pronounced loops lie in the corotation region. The enlargement at right shows orbits of the x_1 family (vertical ellipses) and of the x_2 family (horizontal ellipses). b, Each curve shows the trace in the longitude-velocity diagram of one of the closed orbits of a, when the orbit is viewed along the dotted lines. The wildly snaking curves in the left panel are the traces of orbits that lie in the corotation region. In Fig. 4 this region contains little gas. The right panel shows the innermost orbits on an enlarged scale. It demonstrates that the trace of the cusped orbit in the (l, v) plane resembles a parallelogram, similar to the densest gas in Fig. 4b and c.

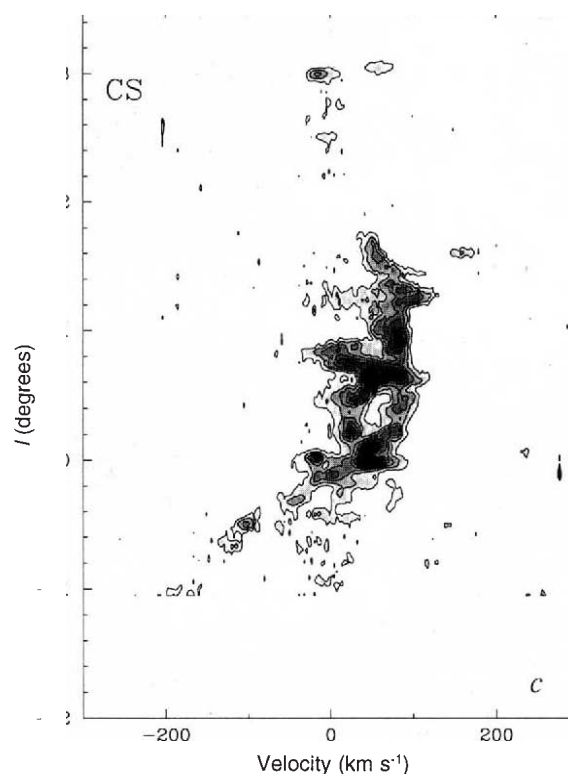
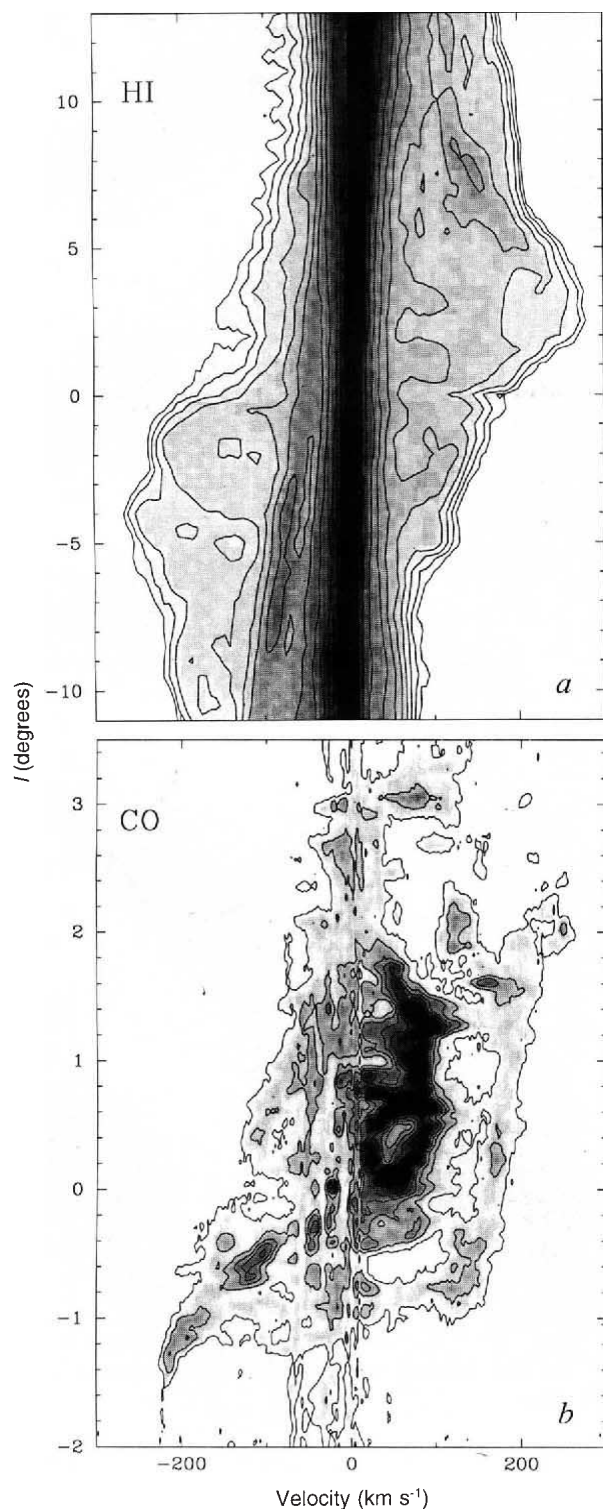
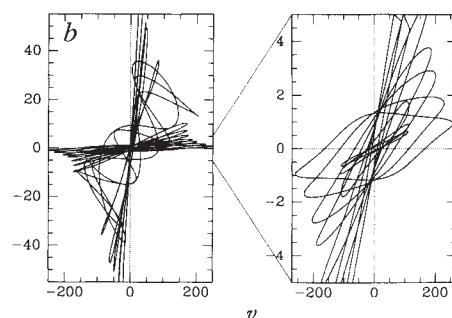
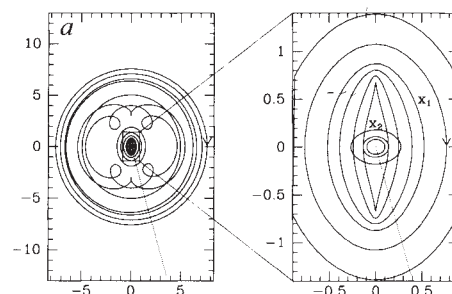


FIG. 4 a, Contours of 21-cm line emission of atomic hydrogen within 6° of the longitude-velocity (l, v) plane⁸⁴. Velocity is relative to the local standard of rest (LSR). The envelope of maximum velocity rises smoothly to a peak near $\pm 3^\circ$ (450 pc). These peaks are believed to be a consequence of the inner Lindblad resonance (ILR) of the galactic bar. The dark band across the middle of the figure is due primarily to atomic hydrogen in the disk between the Sun and the bar. The contours begin at 0.125 K and increase logarithmically, each step being a factor of 2 larger than the previous one. b, Contours of line emission from molecules of CO in the (l, v) plane; the longitude extent is the inner $5\frac{1}{2}^\circ$ degrees of panel a. The narrow bands running across the middle are from foreground gas in the disk. A similar map of the CS emission (c), that traces the densest gas, does not show these bands. The emission at high velocities is associated with a dense molecular disk within 300 pc of the centre. Individual features in this emission are much broader in velocity than any seen in the emission from the disk. The contours in the CO and CS maps are in 10 linear steps from 1 K to 28 K and from 0.2 K to 1.8 K respectively.

If a cloud of gas is released into a potential, the cloud will shear and then settle onto closed orbits, because collisions between parcels of gas tend to dissipate energy of libration around these orbits. Once the gas has settled onto closed orbits, dissipative processes will cause it to drift slowly along a sequence of such orbits to ever lower energies. Consequently, although pressure forces are important in collimating gas onto closed orbits, they are unimportant during the long periods in which the gas drifts from one closed orbit to the next. If the galactic potential were oblate and axisymmetric, the relevant closed orbits would be circular orbits in the equatorial plane of the potential, and a disk of gas that had settled to such orbits would

conform to the conventional picture of the galactic disk.

If the potential is nonspherical, the picture is complicated in two respects. First, the stable closed orbits may at some energy become self-intersecting and second, at a given energy these orbits may form more than one sequence. These phenomena are illustrated in Fig. 5a, which shows three of the key orbit families in the potential of a rotating bar. On all the orbits shown, stars circulate in the same sense as the pattern of the potential. On orbits of the outermost family, stars circulate more slowly than the potential, and conversely for the inner two families. The radius at which the outer orbits develop loops and become self-intersecting is called the 'outer Lindblad resonance'

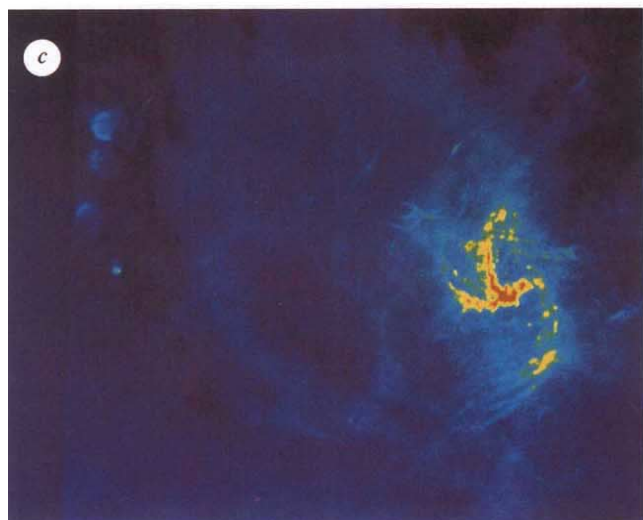
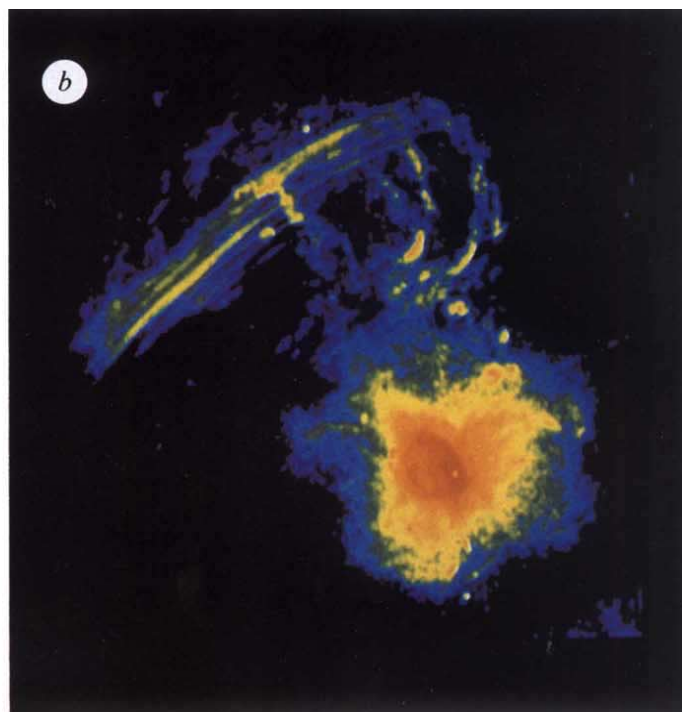
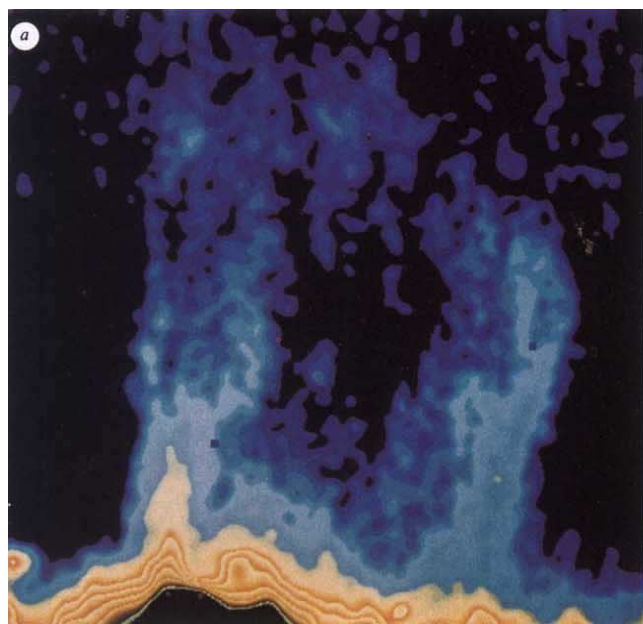
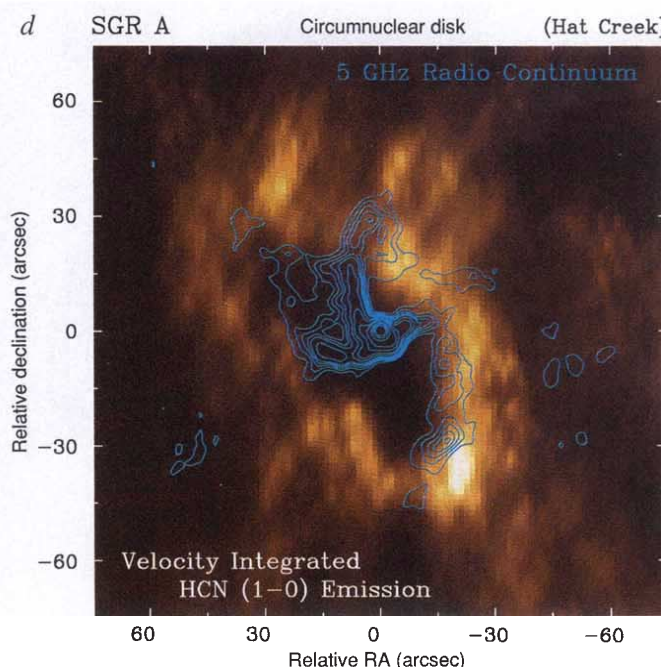


FIG. 6 *a*, An Ω -shaped plume extending about 140 pc (1°) above the plane seen in 11-GHz continuum emission³⁴. The nucleus lies below the bottom of the picture. *b*, A square 70 pc (0.5°) on a side mapped at 1.4 GHz shows a bright knot of emission at lower right known as Sgr A³³. The yellow dot in the red region is Sgr A*, the compact nonthermal source that is believed to mark the nucleus. At upper left is a network of lace-like filaments, which is believed to be organized by magnetic fields. The plane runs roughly diagonally from top left to bottom right. East and positive values of longitude l are to the left, west to the right. *c*, Sgr A, ~ 10 pc across, is resolved into Sgr A East, a nonthermal radio source presumably powered by supernovae (in blue), and Sgr A West, a spiral-like pattern of thermal ionized gas that appears to be falling into the very centre of the Galaxy. *d*, A 2 pc ($1'$) radius rotating ring-like distribution of molecular gas mapped



in HCN line emission⁴². Superimposed on this figure is the spiral pattern of Sgr A West showing how the two are spatially related. The plane again runs roughly from top left to bottom right.

(OLR), and the mean radius of the smallest orbit of the middle family to be self-intersecting is called the 'inner Lindblad resonance' (ILR).

Imagine that gas is being fed onto the outermost orbit in Fig. 5a. As dissipative processes within the gas radiate energy, each cloulet gradually drifts inwards until it approaches the OLR. There the rates of collisions and shocks within the gas increase markedly as the structure of the closed orbits to which collisions drive the gas becomes more elaborate. But the inward drift slows slightly as the extra dissipative losses are more than compensated for by the energy and angular momentum which the bar pumps in at this point—recall that the bar is here going around faster than the gas, so it has a tendency to spin the gas up as well as a tendency to cause it to radiate vigorously.

One of the important open questions is whether radiative losses overwhelm the work done by the bar when the gas reaches the outermost self-intersecting orbit. If so, the gas will plunge rapidly inwards and settle on the middle set of closed elliptical orbits, called the ' x_1 family'²¹. But even if not, mass lost from evolved bulge stars, expected to be of the order of 0.2 solar masses $M_\odot$ per year^{22,23}, will find itself on self-intersecting orbits, and will fall toward the centre. Ultimately, quiescent inward streaming will resume until the ILR is reached, at which point radiative losses again stimulate an abrupt switch to a new orbit family of smaller radius, the innermost set in Fig. 5a, known as the ' x_2 family'. In this way, gas is continually being fed ever closer to the centre from larger radii.

To compare models such as those shown in Fig. 5a with the observations shown in Fig. 4, it is necessary to make a transformation from what one would observe looking down on the plane to the (l, v) plot observed from our position in the plane. Figure 5b shows how the system of Fig. 5a would appear to an observer located at the Sun if the line-of-sight makes an angle of 16° with the major axis of the bar: this is the configuration that gives the best agreement¹⁷ with the data. In particular, the envelope of velocities observed in the H I and CO are well matched to different classes of x_1 orbits, and the giant Galactic Centre molecular clouds Sgr A and Sgr B2 lie, as expected, on x_2 orbits. Furthermore, the region between the inner and outer Lindblad resonances is observed to be fairly clear of gas, and the accumulation of shocked gas outside the OLR can be identified with the well known 'molecular-ring' in the galactic disk.

We see that there is both good photometric and dynamical evidence for a bar at the Galactic Centre, that the orientation of the bar is fairly tightly constrained by the H I and CO observations, and that the dynamics of the bar explain the large-scale kinematics of both the atomic and molecular gas at the centre. The region of the Galaxy commonly referred to as the bulge is thus synonymous with the bar. Any gas flowing inward across the OLR will naturally be driven by the bar into the region within a few hundred parsecs of the nucleus, the region of the molecular disk. In any event, the bar will feed gas ejected by its stars to the nucleus. Let us now look at the molecular disk in detail and try to understand why the gas seems to be so different there.

The inner 300 pc: high gas pressure

The molecular gas shown in Fig. 4b and c lies within 300 pc of the nucleus and has a mass of $\sim 10^8 M_\odot$. This is $\sim 5\%$ of the total molecular mass of the Galaxy, concentrated in $\sim 0.04\%$ of its surface area; the surface density of molecular gas in the centre is extraordinary. The individual clouds, identifiable as discrete features in Fig. 4b and c, are very different from the clouds observed in the disk; they are, on average, much denser and their velocity widths are much larger.

In the CO map of Fig. 4b, several narrow bands of foreground clouds can be seen to cross in front of the broad emission from the centre. The bands in the disk are narrow because individual clouds have both smaller linewidths and smaller random velocities than clouds in the bulge. Because the CS map shows

a striking absence of these narrow bands owing to intervening disk gas, we infer that CS is confined to the bulge. Given that the CS molecule requires higher densities for its excitation than CO, it follows that the gas density ρ in the bulge is about an order of magnitude greater than in the disk. Moreover, the CO and CS lines from individual features in the bulge are broad; the corresponding Doppler velocities are of order $\sigma_v \approx 15 \text{ km s}^{-1}$. It is not adequately understood how molecular clouds in both the disk and the bulge manage to sustain internal motions in excess of the sound speed ($\leq 1 \text{ km s}^{-1}$), but we suspect that strong magnetic fields within the clouds ensure that the Alfvén speed is at least comparable to σ_v . (A field of order $130 \mu\text{G}$ is required; by comparison, the local diffuse gas has a field $\sim 5 \mu\text{G}$ (refs 24, 25). If the Alfvén speed is of order σ_v , it follows that the pressure of the molecular gas in the bulge, $\sim \rho \sigma_v^2$, is about three orders of magnitude higher than that of the disk gas.

Why should this be so? An answer comes from X-ray observations by the Ginga satellite of iron atoms in the bulge that are so highly ionized that they retain only two electrons²⁶. These suggest that the bulge is permeated by very hot gas with temperatures near 10^8 K (hotter than the centre of the Sun) and densities $\sim 0.03\text{--}0.06$ hydrogen nuclei per cubic centimetre, corresponding to a pressure nearly 1,000 times higher than that measured in the solar vicinity. Because the hot gas is pervasive, cooler gas embedded in it must be under the same high pressure. The X-ray-emitting gas is so hot that it must be rushing out of the bulge as a galactic 'wind', unless it is somehow confined by magnetic fields. But even if magnetic fields could confine the wind, the hot gas pressure would be expected to build until it was episodically released as in a pressure cooker. The mass loss rate from the bulge wind depends on the radius r_s of the sonic point for the bulge, and is equal to $0.25 (r_s/200 \text{ pc})^2 M_\odot \text{ yr}^{-1}$. This large mass loss rate is probably enough to compensate for the inflow driven by the bar. In fact, little of the gas driven inward need ultimately fall into the nucleus of the Galaxy. The wind is also likely to be important in carrying off much of the angular momentum of the gas that falls in from more distant regions of the Galaxy. To get within 1 pc of the nucleus from a distance of 1 kpc, an average parcel of gas must have reduced its angular momentum by a factor of at least 10^3 .

One might expect that with such a large concentration of high-density, highly pressured molecules, the star formation rate at the centre would be significantly higher than that in the disk. What is particularly striking about the molecular gas at the centre, however, is the relative paucity of present-day star formation^{27,28}. This could be a direct consequence of the high internal velocity dispersion in the molecular clouds. On account of it, the Jeans mass of the clouds exceeds $10^6 M_\odot$; to be gravitationally bound, clouds have to be extremely massive²⁹. The most spectacular star-forming clouds such as Sgr A and Sgr B2 have masses of this order. Furthermore, the large magnetic fields likely in the Galactic Centre clouds would have to diffuse out before the clouds could collapse. Ambipolar diffusion times may be as high as 10^8 yr for these clouds³⁰. Moreover, to be gravitationally bound a cloud must be dense as well as massive or it will not be able to withstand the strong galactic tidal field near the centre. Unless the circular speed of the Galaxy's potential declines markedly towards the centre, only clouds dense enough to be strong CS emitters could survive shredding by the tidal field^{27,31}.

The central 30 pc

The bar does not greatly affect the orbits of stars and clouds that are confined by the Galaxy's potential to $r \leq 30 \text{ pc}$; the smallest orbits in Fig. 5a form a neatly nested sequence of only slightly eccentric ellipses. Hence one might expect the orbiting gas to become more and more quiescent the farther it recedes from the stimulus of the ILR. Observations, however, decisively confound this expectation. Far from settling to orderly streaming around the nucleus, the region $r < 30 \text{ pc}$ abounds in exotic

phenomena. It is not yet clear why the activity becomes so much more pronounced as the nucleus is approached, although there are several possible explanations. (1) A smaller, more rapidly tumbling bar may sit at the core of the kiloparsec-scale bar discussed above³². (2) As one approaches the nucleus, the energy input from an explosive event required significantly to perturb the flow of gas dwindles. Hence the disorder observed within ~ 30 pc may simply reflect the increasing importance of individual supernova explosions. (3) The magnetic field in the disk and its coupling to the coronal gas above and below the plane may grow in strength to the point at which it destabilizes the planar flow.

Support for this last conjecture is provided by radio continuum observations of filaments which form an arc-like structure near $l = 0.1^\circ$ (14 pc) (Fig. 6b)³³. This structure extends over 0.3° (45 pc) but is part of an even more extensive 1° (140 pc) Ω -shaped lobe that appears to be perpendicular to the plane³⁴ and is reminiscent of the prominences and flares that protrude from the surface of the Sun (Fig. 6a). The intricate and narrow ($\sim 5''$; 0.2 pc) filaments of Fig. 6b are strikingly linear and mostly run parallel to one another. Their steep radio spectra and polarization properties suggest that they emit synchrotron radiation, and their morphology suggests a poloidal component of the galactic magnetic field. This field component would connect gas in the plane to the hotter gas above and below and is likely both to generate instabilities³⁵ and to bleed angular momentum from the disk, possibly leading to the formation of a jet^{36,37}. Such a jet may have been detected³⁸, but more work needs to be done to confirm such an interpretation.

Further support for the suggestion that the magnetic field has a substantial poloidal component is provided by observations of γ -rays. The presence of cosmic rays can be inferred from high-energy γ -rays because these are produced when cosmic rays pass through interstellar gas. The strong nonthermal radio continuum emission from the centre shows that there is a large flux of synchrotron electrons at the centre, and if these are accompanied by relativistic protons, the inner 300 pc of the Galaxy should be a powerful γ -ray source on account of the large column density of molecular gas there. But the COS-B satellite detected no γ -rays from the centre above the background from the disk³⁹. A plausible explanation is that a poloidal magnetic field draws cosmic-ray protons out of the plane before many have interacted with molecular gas, which is confined to a very thin layer (~ 40 pc thick).

Sagittarius A

The central 10 pc is dominated by the radio continuum source known as Sgr A (Fig. 6b), and a dense star cluster observed in the infrared. Sgr A comprises two components, Sgr A East and Sgr A West (Fig. 6c), and a $\sim 10^5 M_\odot$ cloud of molecular gas. Sgr A East has an extent of $3'$ (8 pc). Its nonthermal spectrum⁴⁰ indicates that it is probably powered by supernovae. Sgr A West is smaller (2 pc), its radiation is largely thermal, and it sits within a highly turbulent circumnuclear disk of molecules that extends out to a distance of 7 pc from the nucleus⁴¹. The inner edge of the disk is delineated by a ring of emission observed through its emission of HCN⁴² (see Fig. 6d). Within the molecular gas disk is a cavity 2 pc in radius with a mean gas density about an order of magnitude lower than its surroundings⁴³; Sgr A West sits in this cavity. What makes Sgr A West particularly striking is the spiral-like pattern of emission, with Sgr A* near its centroid. Sgr A* is a point-like radio continuum source that many suspect is the nucleus of the Milky Way, and which signals the presence of a black hole. Within $\sim 1''$ of Sgr A* is the centroid of a dense cluster of stars, much denser than anything observed in the galactic disk. The density of stars within 1 pc of the centre of the cluster is $\geq 10^6 M_\odot \text{pc}^{-3}$ (ref. 44); for comparison, the density of stars in the solar vicinity is $0.1 M_\odot \text{pc}^{-3}$. The central question in the study of the nucleus is whether the mass distribution and energetics are dominated by stars, or by a central black

hole. We describe the observations of Sgr A with a view to answering these questions.

Sgr A West contains $\sim 60 M_\odot$ of ionized gas^{45,46} and may be associated with atomic gas having a mass perhaps five times greater⁴³. The motion of the ionized spiral arms has been studied spectroscopically at both radio and infrared wavelengths^{47,48} and found to be highly organized. From the smoothness of the velocity field within the filaments and the way in which it matches up with the velocities of neighbouring clouds, one infers that the arms are probably gas streams tidally stretched out as they fall towards the bottom of the galactic potential^{46,49,50}. Polarization measurements of these streams in the mid-infrared⁵¹, and milligauss field strengths obtained from Zeeman measurements in the circumnuclear molecular material that envelops the streams⁵², suggest that the magnetic field may play a role in their dynamics.

Figure 6d shows how neatly Sgr A West fits inside the cavity of the circumnuclear molecular material. This fit and the overall similarity of the velocities in the ionized and molecular gas suggest that the circumnuclear molecular gas is the reservoir of the gas that feeds into the nucleus at a rate of $\sim 0.03 M_\odot \text{yr}^{-1}$ (ref. 43). How is the circumnuclear molecular disk related to the molecular clouds farther from Sgr A*? Recent interferometric observations of NH_3 emission from the molecular cloud to the south of Sgr A show that some of the neutral gas in the south may be flowing toward the Galactic Centre in the form of a thin streamer^{53,54} and may settle into the circumnuclear molecular disk.

There is now ample evidence that Sgr A East, most probably driven by stellar winds and supernovae, is blasting into the molecular material eastwards of it, compressing and pushing the cooler gas outwards⁵⁵⁻⁵⁷. Similarly, another supernova remnant $3'$ (7 pc) southwest of Sgr A* seems to be blasting into the Sgr A molecular cloud^{58,59}. By distorting them, supernova explosions may lead to the shredding of otherwise stable clouds by the tidal field of the nucleus. Also, the turbulence to which explosions give rise allows angular momentum to diffuse outwards, and gas to drift inwards, deeper into the gravitational potential. Thus, there may be a feedback cycle causing gas to fall towards the centre. The mass inflow rate into the central cavity is small compared to what is lost by the bulge wind. If the bulge wind is in a steady state and is fed by inflow driven by the bar, the mass flowing into the central parsec is small compared with what is brought in from the disk.

The central parsec

Outside the central parsec, stars (other than supernovae) primarily influence nuclear activity through their collectively generated, barred gravitational potential. Within the central parsec, stars play a more direct role because they are so tightly packed that collisions between them, including tidal disruption, may be sufficiently common⁶⁰ to inject significant quantities of gas into the interstellar medium.

The central star cluster⁶¹ is dominated by a number of discrete sources; one of them, IRS16, has traditionally been considered the core of the cluster. IRS 16 is $\sim 1''$ (0.04 pc) east of Sgr A* and has itself been resolved into about 2 dozen compact components⁶². These are intrinsically rather blue, suggesting that the cluster is composed of fairly young stars⁶³⁻⁶⁵. Moreover, using near-infrared imaging spectroscopy, the central star cluster was found to contain about a dozen He I emission-line stars within $12''$ (0.5 pc) of Sgr A*⁴. These stars have been suggested to be either blue supergiants or Wolf-Rayet-like stars with heavy mass loss ($\sim 10^{-5}$ – $10^{-4} M_\odot \text{yr}^{-1}$), and outflow velocities $\sim 1,000 \text{ km s}^{-1}$; their presence has been taken as evidence for recent star formation. The He I stars are known to be at the Galactic Centre because many are coincident with temperature peaks in the dust associated with the spiral of ionized gas flowing into the centre⁶⁶.

The discovery of these He I emission-line stars suggests that a black hole is not needed to explain the luminosity of the Galactic Centre. These stars are both luminous enough and hot enough to account for the total luminosity within 5 pc of the nucleus, estimated to be $0.5\text{--}2.0 \times 10^7 L_{\odot}$ (refs 67, 68). But the nature of the He I stars is a bit of a mystery, as they have rather unusual spectra. They may be remnants of severe tidal encounters or even mergers between normal stars. Consequently, one cannot securely argue that these stars demonstrate that massive star formation has taken place in the nucleus within the past few million years, and that massive star formation continues there. Moreover, star formation within the central parsec requires gas densities higher than $3 \times 10^7 \text{ cm}^{-3}$ to avoid tidal disruption⁶⁰, a value much higher than what is observed at present. Thus, although the He I stars obviate the need for a black hole to account for the energetics of the galactic nucleus, the questions remain, what powers the enigmatic source Sgr A* and does the latter mark the presence of a black hole?

Sgr A*. It is widely believed that the vast luminosity (up to $10^{45} \text{ erg s}^{-1}$) arising from a very compact volume (a light-year or less) in quasars and Seyfert galaxies are powered by black holes with masses of the order of $10^9 M_{\odot}$, and that such black holes should reside in the nuclei of many, perhaps most galaxies as luminous as the Milky Way. Although it seems that the luminosity of the centre can be provided by the central star cluster, there remain compelling reasons for searching for a possibly dormant massive black hole. Lynden-Bell and Rees² suggested that such an object would be marked by a compact nonthermal radio source. In 1974 Balick and Brown⁶⁹ unambiguously detected such a source, Sgr A*.

Sgr A* is a unique radio source in the Milky Way and shares several properties with compact extragalactic radio sources. Although its structure is still not known, it is very compact, with a linear size $\leq 10 \text{ AU}$ ⁷⁰. (1 AU is the radius of the Earth's orbit: $1.5 \times 10^{13} \text{ cm}$). This small size strongly constrains the nature of its energy source. It rules out a star cluster and requires an object of stellar dimensions or less. A reasonable model for Sgr A* is the corona of a black hole that is accreting very subcritically⁷¹, with synchrotron radiation extending to the mid-infrared. Submillimetre observations of Sgr A* suggest the possibility that a circumnuclear accretion disk with a diameter of $\sim 0.1 \text{ pc}$ may have been detected⁷², although the interpretation is not unique. Another model⁷³ of Sgr A* takes into account the latest observational evidence: the radio emission from Sgr A* results from a supermassive black hole accreting from the winds of He I emission-line stars in the central few arcseconds.

Unlike quasars and active galactic nuclei, which seem to require black holes to explain the enormous energies coming from very compact regions, the energy generated by Sgr A* is rather feeble. From the vantage point of an external galaxy, such as M31, the nucleus of the Galaxy would not be considered active; the radio luminosity of Sgr A*, $\sim 10^{35} \text{ erg s}^{-1}$, is small enough to be provided by a star. It has been suggested, for example, that Sgr A* is just a neutron star moving through a dense gas cloud⁷⁴. The compactness of Sgr A* requires that the emitting material be confined by ambient pressure $nT \sim 10^{12} \text{ K cm}^{-3}$. The most plausible source of such an extreme pressure is ram pressure generated by the motion of Sgr A* through the surrounding gas, but the current upper limit to its proper motion (see below) makes this interpretation unlikely. Moreover, if Sgr A* is merely a luminous neutron star, the uniqueness of Sgr A* is surprising given that there should be many neutron stars in the nucleus.

There is considerable circumstantial evidence that the Galactic Centre contains a point mass of $\sim 10^6 M_{\odot}$. The evidence comes primarily from modelling the observed velocity distributions of the stars^{44,61}, and modelling the measured velocities of the ionized streamers with the assumption that the ionized gas is responding only to gravity⁵⁰. Models with an extended stellar distribution together with a point mass of $2\text{--}3 \times 10^6 M_{\odot}$ seem to

fit the data best⁷⁵. The point mass would presumably be coincident with Sgr A*. Star clusters with very sharply peaked mass distributions ($\rho_*(r) \propto r^{-\alpha}$, where $\alpha < -2.7$) or with small core radii could also fit most of the data. Such sharp density distributions are not observed in other clusters, however, and require extraordinary stellar densities. On the other hand, if Sgr A* marks the presence of a $\sim 10^6 M_{\odot}$ black hole, the 1" offset from the IRS16 cluster is an improbable configuration. One would expect the centroids of both sources to be coincident unless Sgr A* marked a black hole of much more moderate mass⁷⁶.

If Sgr A* were a massive black hole, it would settle at the bottom of the galactic potential well and exhibit negligible proper motion. A neutron star, on the other hand, would show a transverse velocity comparable to the velocity dispersion measured for the stars of $\sim 100 \text{ km s}^{-1}$. The existing radio interferometric data are consistent with Sgr A* being stationary in the Galaxy to better than 27 km s^{-1} (ref. 77). There is less than a 5% probability that an object belonging to a population with velocity dispersion 120 km s^{-1} moves so slowly; alternatively, a $20 M_{\odot}$ object in equipartition with solar-mass stars having 120 km s^{-1} dispersion, has about even chance of moving more slowly than 25 km s^{-1} . To have 99% confidence that the mass of the object is greater than $100 M_{\odot}$, which exceeds the mass of any stable star, one must show that $v < 1.7 \text{ km s}^{-1}$. The current limits are tantalizing, but inconclusive. As time goes on, the limits should get tighter.

In the mid-1970s an argument for the existence of a black hole at the Galactic Centre came from the discovery of a γ -ray line at 511 keV (ref. 78) attributed to the annihilation of electrons and positrons from the general direction of the nucleus. The angular resolution of the observations was inadequate to show that the source was definitely located at the nucleus, but the strength and variability^{79,80} of the line suggested that it arose from a compact source. Consideration of the γ - γ production of positrons requires the source to be very compact, $< 5,000 \text{ km}$, consistent with the observations of Sgr A*. One possible source of the positrons is pair production in intense bremsstrahlung emission from $\sim 10^7 \text{ K}$ gas around an accreting black hole with mass $< 10^3 M_{\odot}$ (ref. 81). More recent imaging results from the SIGMA telescope aboard the GRANAT satellite⁸² show, however, that the variable component of the 511 keV line emission originates from a source that lies $0.75''$ (105 pc) from Sgr A*. The dynamical centre of the Galaxy cannot lie so far from Sgr A*, so the γ -ray line source cannot mark the nucleus.

The search for a black hole at the Galactic Centre remains inconclusive at present. Despite the uncertainties, the unique, compact nature of the Sgr A* radio source, and the high velocities of the ionized gas and stars close to the nucleus provide the best evidence for a concentrated mass of up to $3 \times 10^6 M_{\odot}$. Furthermore, if there is a massive black hole in the nucleus, the low transverse velocity suggests that Sgr A* is it. One prediction of the black-hole model is a large outburst of luminosity after an increase of accretion onto the black hole. Given the amount of matter in the vicinity of Sgr A*, an outburst is not an unreasonable expectation, although since its discovery in 1974, the luminosity has not changed by more than a factor of two. If the luminosity increases beyond $10^{38} \text{ erg s}^{-1}$, 1,000 times the present luminosity, the combination of small scale size and energetics would make the black hole model much more compelling. Nevertheless, many astrophysicists remain unconvinced by the current evidence, and perhaps the only point of general agreement is that the nucleus does not contain a black hole as massive as those expected to reside in extinct quasars. $\square$

Leo Blitz is at the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. James Binney is at the University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK. K. Y. Lo is at the University of Illinois, Astronomy Department, 1002 W. Green Street, Urbana, Illinois 61801, USA. John Bally is at the University of Colorado, APAS, Campus Box 391, Boulder, Colorado 80309, USA. Paul T. P. Ho is at the Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

1. Jansky, K. *Proc. IRE* **20**, 1920 (1932).
2. Lynden-Bell, D. & Rees, M. J. *Mon. Not. R. astr. Soc.* **152**, 461–475 (1971).
3. Wollman, E. R., Geballe, T. R., Lacy, J. H., Townes, C. H. & Rank, D. M. *Astrophys. J.* **218**, L103–L107 (1977).
4. Krabbe, A., Genzel, R., Drapatz, S. & Rotaciuc, V. *Astrophys. J.* **382**, L19–L22 (1991).
5. Hauser, M. G. *et al.* NASA photograph G90-03046 (1990).
6. Hubble, E. *Astrophys. J.* **64**, 321–369 (1924).
7. de Vaucouleurs, G., in *IAU Symp. 20: The Galaxy and the Magellanic Clouds* (eds Kerr, F. J. & Rodgers, A. W.) 195–199 (Australian Academy of Science, Sydney 1964).
8. Peters, W. L. *Astrophys. J.* **195**, 617–629 (1975).
9. Liszt, H. S. & Burton, W. B. *Astrophys. J.* **236**, 779–797 (1980).
10. Yuan, C. *Astrophys. J.* **281**, 600–613 (1984).
11. Blitz, L. & Spergel, D. N. *Astrophys. J.* **379**, 631–638 (1991).
12. Matsumoto, T. *et al.* in *The Galactic Center* (eds G. Riegler & R. Blandford) 48–52 (American Institute of Physics, New York, 1982).
13. Weiland, J. *et al.* in *Back to the Galaxy* (ed. Verter, F.) (AIP, New York, in the press).
14. Weinberg, M. D. *Astrophys. J.* **384**, 81–94 (1992).
15. Nakada, Y. *et al.* *Nature* **353**, 140–141 (1992).
16. Whiteelock, P. & Catchpole, R. in *The Center, Bulge and Disk of the Milky Way* (ed. Blitz, L.) 103–110 (Kluwer, Dordrecht, 1992).
17. Binney, J. J., Gerhard, O. E., Stark, A. A., Bally, J. & Uchida, K. I. *Mon. Not. R. astr. Soc.* **252**, 210–218 (1991).
18. Cohen, R. J. & Few, R. W. *Mon. Nat. R. astr. Soc.* **176**, 495–523 (1976).
19. Cohen, R. J. & Dent, W. R. F. in *Surveys of the Southern Galaxy* (eds Burton, W. B. & Israel, F. P.) 159–164 (Reidel, Dordrecht, 1983).
20. Bally, J., Stark, A. A., Wilson, R. W. & Henkel, C. *Astrophys. J. Suppl.* **65**, 13–82 (1987).
21. Contopoulos, G. & Mertzaniades, C. *Astr. Astrophys.* **61**, 477–485 (1977).
22. Kent, S. M., Dame, T. M. & Fazio, G. *Astrophys. J.* **378**, 131–138 (1991).
23. Ciotti, L., D'Ercole, A., Pellegrini, S. & Renzini, A. *Astrophys. J.* **376**, 380–403 (1991).
24. Heiles, C. *Astrophys. J.* **324**, 321–330 (1988).
25. Boulares, A. & Cox, D. P. *Astrophys. J.* **365**, 544–558 (1990).
26. Yamauchi, S. *et al.* *Astrophys. J.* **365**, 532–538 (1991).
27. Güsten, R. in *The Center of the Galaxy* (ed. Morris, M.) 89–106 (Kluwer, Dordrecht, 1989).
28. Morris, M., Yusef-Zadeh, F. & Chance, D. in *Birth and Infancy of Stars* (eds Lucas, R., Omont, A. & Stora, R.) 209 (Elsevier, Amsterdam, 1985).
29. Spergel, D. N. & Blitz, L. *Nature* **357**, 665–667 (1992).
30. Morris, M. in *The Center of the Galaxy* (ed. Morris, M.) 171–177 (Kluwer, Dordrecht, 1989).
31. Stark, A. A., Bally, J., Wilson, R. W. & Pound, M. W. in *The Center of the Galaxy* (ed. Morris, M.) 129–133 (Kluwer, Dordrecht, 1989).
32. Shlossman, I., Frenk, J. & Begelman, M. *Nature* **338**, 45–47 (1989).
33. Yusef-Zadeh, F., Morris, M. & Chance, D. *Nature* **310**, 557–561 (1984).
34. Sofue, Y. & Handa, T. *Nature* **310**, 568–569 (1984).
35. Balbus, S. A. & Hawley, J. F. *Astrophys. J.* **376**, 214–222 (1991).
36. Blandford, R. D. & Payne, D. R. *Mon. Nat. R. Astr. Soc.* **199**, 883–903 (1982).
37. Pudritz, R. E. & Norman, C. A. *Astrophys. J.* **301**, 571–586 (1986).
38. Sofue, Y., Reich, W. & Reich, P. *Astrophys. J.* **341**, L47–L49 (1989).
39. Blitz, L., Bloemen, J. B. G. M., Hermesen, W. & Bania, T. M. *Astrophys. J.* **143**, 267–273 (1985).
40. Pedlar, A. *et al.* *Astrophys. J.* **342**, 769–784 (1989).
41. Genzel, R. in *The Center of the Galaxy* (ed. Morris, M.) 393–405 (Kluwer, Dordrecht, 1989).
42. Güsten, R. *et al.* *Astrophys. J.* **318**, 124–138 (1987).
43. Jackson, J. M. *et al.* *Astrophys. J.* **402**, 173–184 (1993).
44. McGinn, M. T., Sellgren, K., Becklin, E. E. & Hall, D. N. B. *Astrophys. J.* **338**, 824–840 (1989).
45. Ekers, R. D., van Gorkom, J. H., Schwarz, U. J. & Goss, W. M. *Astr. Astrophys.* **122**, 143–150 (1983).
46. Lo, K. Y. & Claussen, M. *Nature* **306**, 647–651 (1983).
47. Lacy, J. H., Geballe, T. R. & Hollenbach, D. J. *Astrophys. J.* **262**, 120–134 (1980).
48. Roberts, D. A., Goss, W. M. & van Gorkom, J. H. *Astrophys. J.* **366**, L15–L18 (1991).
49. Quinn, P. J. & Sussman, G. J. *Astrophys. J.* **288**, 377–384 (1985).
50. Serabyn, E. & Lacy, J. H. *Astrophys. J.* **293**, 445–458 (1985).
51. Atkin, D. K., Gezari, D., Smith, C. H. McCaughrean, M. & Roche, P. F. *Astrophys. J.* **380**, 419–428 (1991).
52. Killeen, N. E. B., Lo, K. Y. & Crutcher, R. *Astrophys. J.* **385**, 585–603 (1992).
53. Ho, P. T. P. *et al.* *Nature* **350**, 309–312 (1991).
54. Okumura, S. K. *et al.* *Astrophys. J.* **378**, 127–130 (1991).
55. Mezger, P. G. *et al.* *Astrophys.* **209**, 337–348.
56. Zylka, R., Mezger, P. G. & Wink, J. *Astr. Astrophys.* **234**, 133–146 (1990).
57. Okumura, S. K. *et al.* *Astrophys. J.* **347**, 240–250 (1989).
58. Ho, P. T. P., Jackson, J. M., Barrett, A. H. & Armstrong, J. T. *Astrophys. J.* **288**, 575–579 (1985).
59. Szczepanski, J. C., Ho, P. T. P. & Güsten, R. in *Skylines, Proc. 3rd Haystack Observatory Mtg* (eds Haschick, A. D. & Ho, P. T. P.) 143–151 (Astr. Soc. Pacif., San Francisco, 1990).
60. Phinney, E. S. in *The Center of the Galaxy* (ed. Morris, M.) 543–553 (Kluwer, Dordrecht, 1989).
61. Sellgren, K., McGinn, M. T., Becklin, E. E. & Hall, D. N. B. *Astrophys. J.* **359**, 112–120 (1990).
62. Eckart, A. *et al.* *Nature* **355**, 526–528 (1992).
63. Allen, D. A., Hyland, A. R. & Hillier, D. J. *Mon. Nat. R. astr. Soc.* **244**, 706–713 (1990).
64. Adams, D. J. *et al.* *Astrophys. J.* **327**, L65–L68 (1988).
65. Rieke, G. H. & Lebofsky, M. J. in *The Galactic Center* (eds Riegler, G. R. & Blandford, R. D.) 194–302 (AIP, New York, 1982).
66. Gezari, D. Y. in *The Center, Bulge and Disk of the Milky Way* (ed. Blitz, L.) 23–46 (Kluwer, Dordrecht, 1992).
67. Becklin, E. E., Gatley, I. & Werner, M. W. *Astrophys. J.* **258**, 135–142 (1982).
68. Werner, M. & Davidson, J. *IAU Symp. No. 136, The Center of the Galaxy* (ed. Morris, M.) 423–436 (Reidel, Dordrecht, 1989).
69. Balick, B. & Brown, R. L. *Astrophys. J.* **194**, 265–270 (1974).
70. Lo, K. Y., Backer, D., Kellerman, K., Moran, J. M. & Reid, M. *Nature* **315**, 124–125 (1985).
71. Rees, M. J. in *The Galactic Center* (eds Riegler, G. R. & Blandford, R. D.) 166–176 (AIP, New York, 1982).
72. Zylka, R., Mezger, P. G. & Lesch, L. *Astr. Astrophys.* **261**, 119–121 (1992).
73. Melia, F. *Astrophys. J.* **387**, L125–L28 (1992).
74. Reynolds, S. P. & McKee, C. F. *Astrophys. J.* **239**, 893–897 (1980).
75. Genzel, R. & Townes, C. H. A. *Rev. Astr. Astrophys.* **25**, 377–423 (1987).
76. Sanders, R. H. *Nature* **359**, 131–132 (1992).
77. Backer, D. & Sramek, R. *Abstr. URSI Mtg* (January 1993).
78. Haymes, R. C. *et al.* *Astrophys. J.* **201**, 593–602 (1975).
79. Riegler, G. R., Ling, J. C., Mahoney, W. A., Wheaton, W. A. & Jacobsen, A. S. *Astrophys. J.* **294**, L13–K15 (1985).
80. Leventhal, M., MacCallum, C. J., Husters, A. F. & Stang, P. D. *Astrophys. J.* **260**, L1–L5 (1982).
81. Lingenfelter, R. & Ramaty, R. in *AIP Proc. No. 83* (eds Riegler, G. & Blandford, R.) 148–158 (American Institute of Physics, New York, 1982).
82. Bouchet, L. *et al.* *Astrophys. J.* **383**, L45–L48 (1991).

ACKNOWLEDGEMENTS. We thank D. Backer for communicating results before publication, and R. Genzel, M. Morris, R. Sanders and C. Townes for comments on the manuscript. This work is partially funded by the U.S. NSF.

LETTERS TO NATURE

Ionospheric signatures of pulsed reconnection at the Earth's magnetopause

M. Lockwood*, W. F. Denig†, A. D. Farmer‡, V. N. Davda*, S. W. H. Cowley§ & H. Lühr||

* Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK

† Phillips Laboratory, PL/GPSG, Hanscom AFB, Bedford, Massachusetts 01731-5000, USA

‡ Atmospheric Physics Laboratory, University College London, London W1P 7PP, UK

§ Blackett Laboratory, Imperial College, London SW7 2BZ, UK

|| Institut für Geophysik und Meteorologie, Technische Universität Braunschweig, W-3300 Braunschweig, Germany

THE plasma precipitating into the Earth's dayside auroral atmosphere has characteristics which show that it originates from the shocked solar-wind plasma of the magnetosheath^{1,2}. The particles of the magnetosheath plasma precipitate down a funnel-shaped region (cusp) of open field lines resulting from reconnection of the geomagnetic field with the interplanetary magnetic field³. Although the cusp has long been considered a well defined spatial structure maintained by continuous reconnection, it has recently been suggested^{4–6} that reconnection instead may take place in a series of discontinuous events; this is the 'pulsating cusp model'.

Here we present coordinated radar and satellite observations of a series of discrete, poleward-moving plasma structures that are consistent with the pulsating-cusp model.

Figure 1 shows schematically a noon-midnight cross-section of the Earth's magnetosphere, showing the cusps (C). The dashed line is the magnetopause, the current-carrying boundary between the magnetosphere and the shocked solar wind plasma in the magnetosheath (MS). The interplanetary magnetic field (IMF) is shown pointing southward, allowing it to reconnect with the northward-pointing 'closed' geomagnetic field lines (c) at a neutral line (X) in the subsolar magnetopause. This generates 'open' field lines (o) which thread the boundary and connect the magnetosphere with interplanetary space. The open field lines convect polewards under the joint action of magnetic 'tension' and the solar wind flow. If the rate at which field lines are reconnected exceeds the rate at which they convect polewards, the open/closed boundary migrates toward the Equator. Magnetosheath plasma can enter the magnetosphere along the open field lines. The large time-of-flight of the ions disperses them according to their energy³: highest-energy ions arrive first in the atmosphere and are seen at the lowest latitudes. This is because more recently reconnected field lines are always closer to the open/closed field line boundary which lies at the equatorward edge of the cusp⁴.

If the reconnection were to take place at a steady rate, the ion energy would show a continuous latitudinal dispersion. However, the pulsating cusp model predicts that the dispersion

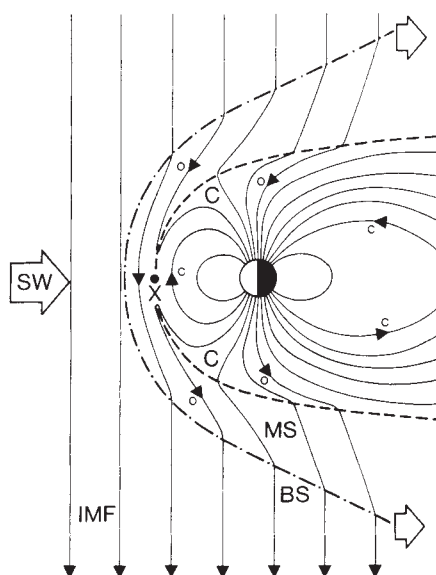


FIG. 1 Schematic of the dayside magnetosphere for southward interplanetary magnetic field (IMF). The super-Alfvénic solar wind (SW) forms a bow shock (BS), behind which lies the shocked plasma of the magnetosheath (MS). Reconnection takes place at X, converting closed field lines, c, into open field lines, o, which form a rotational field discontinuity where they cross the magnetopause (dashed line). This results in the open field lines convecting polewards under the magnetic 'tension' force, and in doing so they accelerate the magnetosheath ions that flow along them and cross the magnetopause^{20,23}. As they straighten, open field lines continue to convect polewards because they are embedded in the solar wind flow, but the ion acceleration decreases (and turns to a deceleration poleward of the magnetic cusp). If the IMF has a significant dawn/duskward B_y component (out of the plane of the diagram), magnetic tension also causes east/west motion on the most recently reconnected field lines^{6,7,19}. The injected magnetosheath plasma precipitates down the convecting open field lines, into the ionosphere in the cusps (C). The 'pulsating cusp' model^{6,7,10} uses the fact that the spectra of particles arriving at the ionosphere on a given field line depend upon the time elapsed since that field line was reconnected. Ions with the lowest energy (E) observed on any one field line have the longest time-of-flight, t , and were therefore the first injected: they crossed the magnetopause at X, a distance d from the satellite ($t = d(2E/m_i)^{-1/2}$, where m_i is the ion mass). That minimum energy falls with time elapsed since reconnection. Hence two pulses of reconnection, with an interval δt between them, cause a step in the minimum energy from E_1 to E_2 where $\delta t = (t_2 - t_1) = d(m_i/2)^{1/2}(E_2^{-1/2} - E_1^{-1/2})$.

FIG. 3 Spectrograms showing the differential number flux (in $\text{cm}^{-2} \text{s}^{-1} \text{sr}^{-1} \text{eV}^{-1}$) of precipitating electrons (top) and ions (bottom) as a function of energy and time, as observed by DMSP-F10 during the pass described in Fig. 2. The decrease of ion energies with latitude at all times (and the observed convection away from the Sun) indicates that the IMF points southward; inverse-slope, V-shaped or dispersionless signatures are expected for northward IMF^{3,6,14}. The arrows mark the boundaries of the 'cusp' (between larger arrows), 'cleft' and 'mantle' precipitations (equatorward and poleward of the cusp, respectively). The cusp-cleft boundary is defined to be where the mean ion energy falls below 3 keV (refs 1, 12), and the cusp-mantle boundary where the ion energy flux falls below $10^{10} \text{ eV cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$. The electron energy flux in the cusp averages $3 \times 10^{10} \text{ eV cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$, which is a factor of two lower than some definitions, under which this region would be classed as 'probably cusp'. However, this category shows the same statistical occurrence as 'cusp' (with electron energy flux exceeding $6 \times 10^{10} \text{ eV cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$) (ref. 12). In the 'cleft', ions of energy up to a maximum of 5.5 keV (i.e. velocity of $1,000 \text{ km s}^{-1}$) are observed, revealing acceleration to above magnetosheath energies. This is the same as the peak field-aligned velocity of the D-shaped

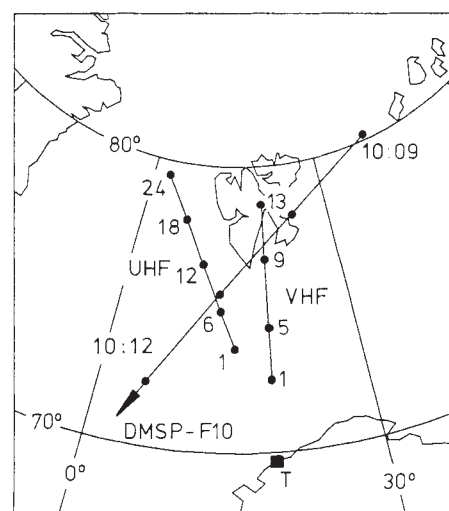
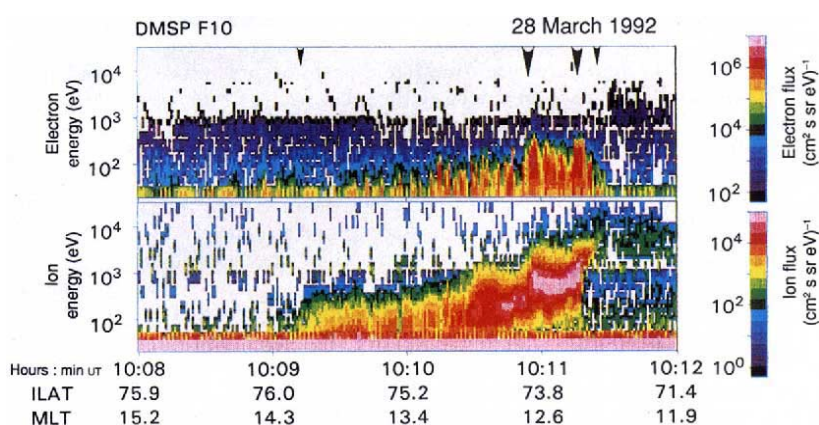


FIG. 2 Map (geographic coordinates) of the EISCAT UHF and VHF radar beams for the SP-UK-CONV experiment, used to study the ionized upper atmosphere in the F-region ionosphere on 28 March 1992: the points show centres of the radar range gates. The altitude of the UHF radar beam varies from 197 km for gate 1 to 595 km for gate 25. Also shown is the path of the DMSP-F10 satellite between 10:09 and 10:12 UT. The satellite passed at 800 km altitude over the EISCAT field-of-view, observing the same field line as the UHF radar at 10:11 UT. In magnetic coordinates, both were then at 12:55 magnetic local time (MLT) and an invariant (magnetic) latitude of $\Lambda = 74^\circ$. Invariant latitude is a way of defining a magnetic field line by the motion of energetic charged particles along it and is the magnetic latitude of the point where the field line intersects the Earth. MLT is defined in terms of the point where the field line would cut the ecliptic plane and the angle that point subtends with the Sun at the Earth's centre. That angle is zero at MLT=12 hr and cusp precipitation is generally found at MLT between 10:30 and 13:30 (ref. 12). Magnetic coordinates of both the radar range gates and the satellite position were determined using the IGRF (International Geomagnetic Reference Field) model of the Earth's magnetic field.



distributions of the accelerated magnetosheath ions at the magnetopause, as both predicted theoretically²³ and, recently, observed²⁴⁻²⁶. This similarity indicates that the 'cleft' is, for this case at least, on the most-recently opened field lines: this is the only place where these highest-energy ions will be observed because of their low time-of-flight.

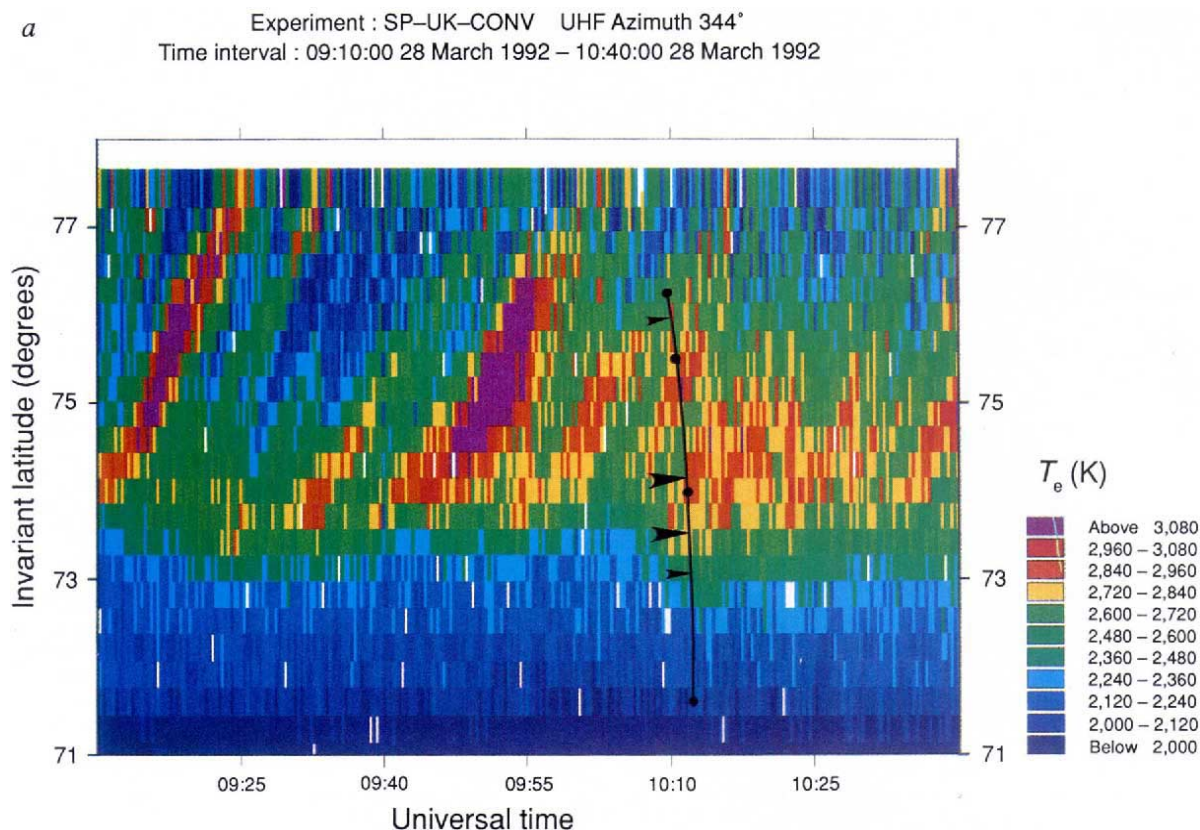


FIG. 4 The electron temperatures observed by the EISCAT UHF radar as a function of invariant latitude (Λ) and time for (a) extended and (b) short periods around the satellite pass. The Λ of the F10 satellite is given by the black line; dots are time markers 1 min apart; the outer small arrows are the edges of the observed magnetosheath-like plasma and the inner, large arrows mark the location of the two major step features in the ion dispersion

signature (Fig. 3). At the time that magnetosheath-like ('mantle') precipitation is first observed, the satellite is at a MLT 1.6 hr further into the afternoon sector than the UHF radar beam at the same Λ (Δ MLT = +96 min), but at the edges of the 'cusp' (the larger arrows) Δ MLT is just +9 min and –6 min and at the equatorward edge of the 'cleft', it is –9 min. Hence for the cusp and cleft regions the satellite–radar conjunction is very close. In *b* we also

signature will show discontinuities if the reconnection is pulsed, rather than steady^{6,7}. The ions appearing at the foot of the most recently reconnected field lines would be more energetic than the corresponding ions on field lines that were opened by a previous pulse of reconnection. Between these two regions, a step in the ion energy will be observed, corresponding to the period of little or no reconnection between the pulses. The steps will drift polewards with the convecting field lines.

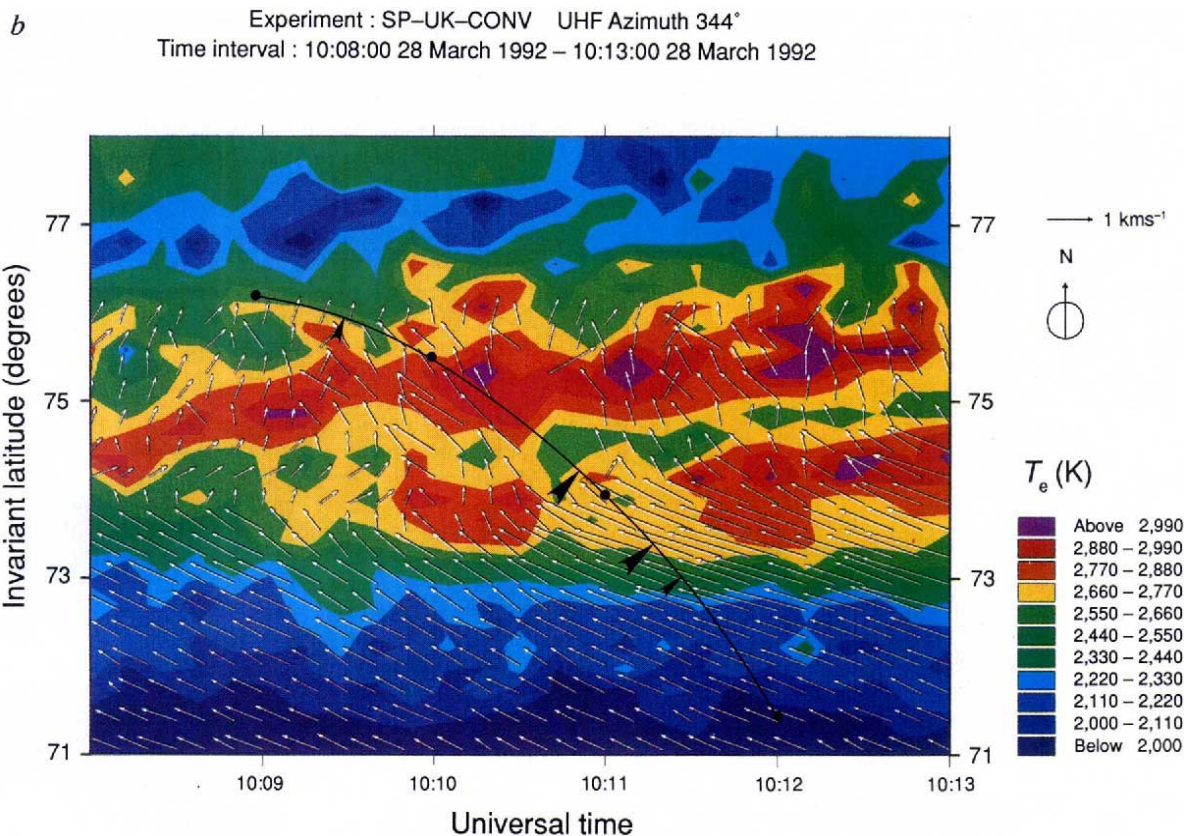
Such 'staircase' signatures are common in cusp ion precipitation^{8,9} and have been used to calculate the reconnection rate as a function of time¹⁰. Hence the pulsating cusp model provides a powerful tool for studying how reconnection proceeds at the dayside magnetopause. However, there has been debate as to the validity of the model^{11–13} with some scientists favouring the original explanation in terms of steady-state spatial structure^{8,14}, dividing the precipitating plasma into 'cleft', 'cusp' and 'mantle' regions^{1,12}.

Ambiguity between spatial and temporal variations is a common problem in interpreting any sequence of data from an orbiting satellite. Ground-based observations, on the other hand, offer an opportunity for remote sensing of the plasma in a given region over a prolonged period. Such measurements distinguish temporal events from spatial structure, but suffer from lower resolution. In the case of the cusp, we can detect transient features from the ground, but cannot unambiguously define them as being due to the cusp precipitation^{15,16}. Here we report combined observations by the EISCAT radars and the DMSP-

F10 satellite which tell us about the spatial and temporal behaviour of the cusp. A map of the observations, made on 28 March 1992, is given in Fig. 2.

Figure 3 shows the energy–time spectrograms of the ions and electrons observed by the F10 satellite; magnetosheath-like plasma was observed between 10:09:15 and 10:11:30 UT. The most striking feature of these data is the ion-energy dispersion which showed discrete steps in the ion spectrum at 10:10:28, 10:10:54 and 10:11:15 UT. The electron fluxes were irregular, with peaks in the precipitating energy flux at the second and third of the steps which bound the 'cusp' precipitation: the regions polewards and equatorwards would usually be termed 'mantle' and 'cleft', respectively^{1,12}.

Figure 4 shows the electron temperature observed by the EISCAT ultra-high-frequency (UHF) radar as a function of invariant latitude, Λ (see Fig. 2), and time. The black line shows the Λ of the satellite for 10:09–10:12 UT, the arrows corresponding to those in Fig. 3. The large arrows mark the locations of two steps bounding the cusp, which were observed very near the UHF beam. The radar reveals a number of poleward-moving regions in which the electron temperature is elevated by a factor of ~ 1.5 . The size, amplitude and repeat period of these events vary and those during the satellite pass are weaker than those seen just a few minutes earlier. These features are reminiscent of those observed in 630-nm dayside auroral emissions at winter solstice¹⁵. We expect heating of the ambient ionospheric electron gas in response to the precipitation of solar wind plasma¹⁷, and the plasma density data (not shown) reveal that weak density



show 10-s flow vectors, derived by combining the line-of-sight velocities from the UHF and VHF radars (where both are available). The satellite measured the component of the drift perpendicular to its orbit path and at all times for which a comparison could be made, this was the same (to within 10%) as that derived from the radar data. The UHF beam elevation is 20°, but we know altitude variations are not a major factor because (1)

the same features were seen at the same time and Λ at greater altitudes by the VHF radar which employed an elevation of 30°; (2) low electron temperatures ($T_e < 2,200$ K) were observed polewards of the events; and (3) the features drifted polewards with the observed poleward component of the plasma convection. Hence the events are poleward-moving latitudinal structures and not rising, extensive, thin layers.

increases emerge within the events late in their lifetime, at the highest latitudes. This indicates that the events are caused by the low-energy particle precipitation and suggests that they may be involved in the production of polar cap density 'patches'¹⁸. The radar data alone cannot tell us that these features are indeed the cusp. However, Fig. 4*b* shows the data during the satellite overpass at higher time resolution and it can be seen that the southernmost event was coincident with the cusp (between the large arrows). A poleward-moving event is also seen in the mantle region. No temperature event is seen under the cleft precipitation in Fig. 4*b*, but 4*a* shows that this emerges as a third poleward-moving event after 10:14 UT. Using the theory outlined in Fig. 1 (ref. 10), the time interval δt (~ 5.3 min) between the events seen by EISCAT and the step in the lower cut-off of the ion energy seen by F10 (from $E_1 \approx 1.2$ keV to $E_2 \approx 200$ eV) yield a distance from the satellite to the neutral line (X in Fig. 1) of $d \approx 10^5$ km (16 Earth radii R_E). This places the reconnection close to the subsolar magnetopause, consistent with previous studies of both 'steady' reconnection^{19,20} and 'flux transfer events' (FTEs)²¹.

By combining the data from the two radars, convection vectors of 10-s resolution can be derived, with the assumption that the flow does not vary between the two beams. The results are also shown in Fig. 4*b*. At all times the poleward convection speed was ~ 500 m s⁻¹, the same as the speed of poleward drift of each of the heated regions as predicted by the pulsating cusp model. The flow rotates from uniformly northwestwards to northwards close to the poleward edge of the cusp. Hence the cusp was on

field lines which were convecting with similar velocity to the closed field lines equatorward of the cusp. This feature is predicted by the pulsating cusp model: a pulse of reconnection is envisaged as initially producing an equatorward motion of the open/closed boundary, rather than causing a burst of poleward flow (with the boundary static). This can excite continuous, even steady, poleward flow as the boundary subsequently relaxes poleward^{4,6,7}. If the y -component of the interplanetary magnetic field $|B_y|$ were large, the tension force would yield a strong zonal flow burst on the most recently opened flux^{6,7,16}: the moderate northwestward flow observed here in the cusp/cleft region implies that B_y is weakly positive; we cannot confirm this inference because no simultaneous measurements were made of the interplanetary magnetic field.

These results show that the magnetosheath-like plasma precipitation was, in fact, made up of three poleward-moving events from an extended sequence of such features. The satellite intersections with these three events would have been termed 'mantle', 'cusp' and 'cleft' in a spatial interpretation, but the radar shows them to be similar events seen at different phases of their evolution. This finding is supported by the steps in the dispersion signature of the ions. Hence the cusp was not, at this time at least, a steady-state spatial feature, but was one of a sequence of poleward-moving events. This is direct evidence for the pulsating cusp model and that reconnection was taking place in a series of bursts (FTEs) at the subsolar magnetopause. The period between the events which can be resolved in Fig. 4*a* varies between about 2 and 15 min. This is consistent with the

range for the magnetopause signatures which are also widely believed to result from FTEs²². □

Received 29 October; accepted 22 December 1992.

- Newell, P. T. & Meng, C.-I. *J. geophys. Res.* **93**, 14549–14556 (1988).
- Heikkilä, W. J. & Winningham, J. D. *J. geophys. Res.* **76**, 883–891 (1971).
- Reiff, P. H. *et al. J. geophys. Res.* **82**, 479–491 (1977).
- Lockwood, M. & Smith, M. F. *Geophys. Res. Lett.* **16**, 879–882 (1989).
- Smith, M. F. & Lockwood, M. *Geophys. Res. Lett.* **17**, 1069–1072 (1990).
- Smith, M. F. *et al. Planet. Space Sci.* **40**, 1251–1268 (1992).
- Cowley, S. W. H. *et al. in CLUSTER-Dayside Polar Cusp*, ESA SP-330, (ed. Barron, C. I.) 105–112 (European Space Agency Publications, Noordwijk, The Netherlands, 1991).
- Newell, P. T. & Meng, C. I. *Geophys. Res. Lett.* **18**, 1829–1832 (1991).
- Escoubert, P. *et al. Geophys. Res. Lett.* **19**, 1735–1738 (1992).
- Lockwood, M. & Smith, M. F. *J. geophys. Res.* **97**, 14841–14847 (1992).
- Newell, P. T. *Geophys. Res. Lett.* **17**, 303–304 (1990).
- Newell, P. T. & Meng, C.-I. *Geophys. Res. Lett.* **19**, 609–613 (1992).
- Lockwood, M. & Smith, M. F. *Geophys. Res. Lett.* **17**, 305–306 (1990).
- Lundin, R. *et al. in CLUSTER-Dayside Polar Cusp* (ed. Barron, C. I.) ESA SP-330, 83–95 (European Space Agency Publications, Noordwijk, The Netherlands, 1991).
- Sandholt, P. E. *et al. Ann. Geophys.* **10**, 483–497 (1992).
- Lockwood, M. *et al. Planet. Space Sci.* **37**, 1347–1365 (1989).
- Curtis, S. A. *et al. Geophys. Res. Lett.* **9**, 997–1000 (1982).
- Tsunoda, R. T. *Rev. Geophys.* **26**, 719–760 (1988).
- Gosling, J. T. *et al. J. geophys. Res.* **95**, 8073–8084 (1990).
- Paschmann, G. *in Magnetic Reconnection in Space and Laboratory Plasmas*, 114–123 (AGU Monogr. 30, 1984).
- Russell, C. T. *et al. Adv. Space Res.* **5**, 363–369 (1985).
- Lockwood, M. & Wild, M. N. *J. geophys. Res.* (in the press).
- Cowley, S. W. H. *Rev. Geophys.* **20**, 531–565 (1982).
- Smith, M. F. & Rodgers, D. J. *J. geophys. Res.* **95**, 11617–11624 (1991).
- Fuselier, S. A. *et al. Geophys. Res. Lett.* **18**, 139–142 (1991).
- Gosling, J. T. *et al. Geophys. Res. Lett.* **17**, 2245–2248 (1990).

ACKNOWLEDGEMENTS. We thank the director and staff of the EISCAT Scientific Association for their assistance and H. C. Carlson for discussions. The work was supported by the U.K. SERC and U.S. AFOSR.

Photochemical switching of polarization in ferroelectric liquid-crystal films

Tomiki Ikeda*, Takeo Sasaki & Kunihiro Ichimura

PRESTO, JRDC, Research Laboratory of Resources Utilization, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 227, Japan

LIQUID crystals have been used extensively as active media in display devices such as full-colour television screens. These devices are generally based on changes in the arrangement of the liquid-crystal molecules induced by electric fields, which change their optical properties¹. Ferroelectric liquid crystals^{2,3} exhibit spontaneous polarization and therefore show a faster response to changes in the applied field. Switching of this field causes a reversal in the direction of polarization^{2–5}. Here we report polarization switching in ferroelectric liquid crystals driven by a photochemical process. The liquid-crystal films are doped with a photochromic compound which undergoes *trans*–*cis* isomerization on irradiation. Photoisomerization induces a change in the switching potential of the host liquid-crystal film, and thereby causes switching at the irradiated sites. The process is fast, stable, reversible and repeatable, and should be exploitable in device applications.

Many studies have been done on the photoresponse of liquid crystals (LCs), with an aim to developing optical image-recording systems^{6–13}. Laser-addressed LC display devices have been demonstrated, based on the thermal control of LC alignment. In contrast to these 'heat-mode' systems, 'photon-mode' devices involve photochemical control and switching of alignment. Haas *et al.*⁹ reported a change in selective reflectivity of a cholesteric LC mixture of cholesteryl bromide and other cholesteryl derivatives on photoirradiation, owing to a change

in the helical pitch of the cholesteric LC induced by a photochemical reaction of cholesteryl bromide. Photochemical reactions of guest molecules in LC hosts can induce a decrease of the nematic–isotropic phase transition temperature (T_{NI}) of the mixture and cause an isothermal phase transition at the irradiated sites^{10–12}. Although such a photochemically induced phase transition might be exploited in optical image-recording systems, the response time and working temperature are disadvantageous. Time-resolved measurements^{14,15} have revealed that the photochemical phase transition of low-molecular-weight or polymeric LCs takes place on a timescale of 50–200 ms. Furthermore, the temperature range is limited because the lowering of T_{NI} due to the photochemical reaction of the guest is generally small ($\leq 10^\circ\text{C}$). To achieve a quicker response and a wider temperature range for optical image recording systems, a different principle is needed.

The chiral smectic C (SmC^*) phase of ferroelectric LCs (FLCs) exhibits spontaneous polarization (P_s), owing to the presence of polar groups which give the molecules an electric dipole moment. In the FLC I , dipole moments arising from the carbonyl moieties lie perpendicular to the long axis of the FLC molecules. In the SmC^* phase, the FLC molecules are aligned parallel to each other to form a layer with a tilt between the direction of the long axis of the FLC molecule and the normal to the smectic layer^{2,3}. The average direction of the molecular long axis (director, $\mathbf{n}$) is defined in each layer, and owing to the chiral group in the FLC molecules the director adopts a helicoidal structure with a characteristic pitch. This means that

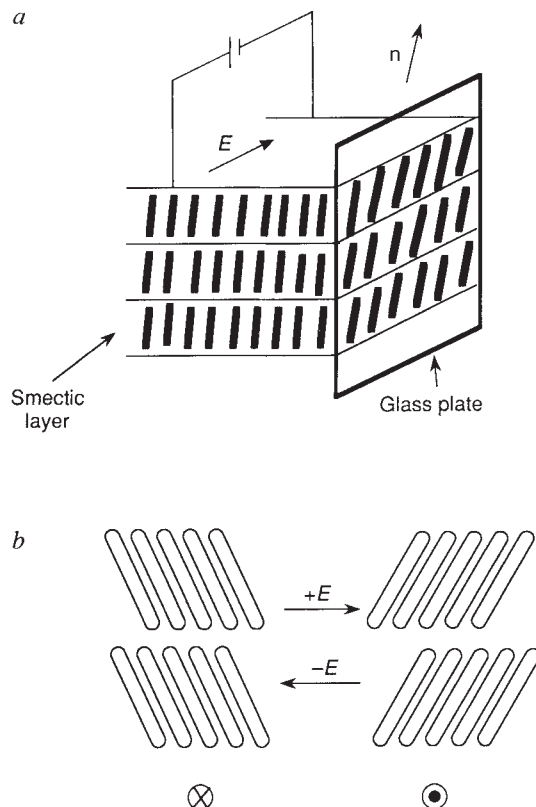


FIG. 1 Orientation of SmC^* layers (a) and direction of spontaneous polarization (b) in the surface-stabilized state. In a, mesogens are aligned parallel to each other to form layers with a tilt between the director ($\mathbf{n}$) and the normal to the smectic layer. The polarization can be made to reverse by applying an external electric field normal to the plane, which simultaneously causes the change in optical property of FLC. When viewed with a pair of polarizers, the right state transmits light while the left state extinguishes, thus producing a contrast between the two states.

* To whom correspondence should be addressed.

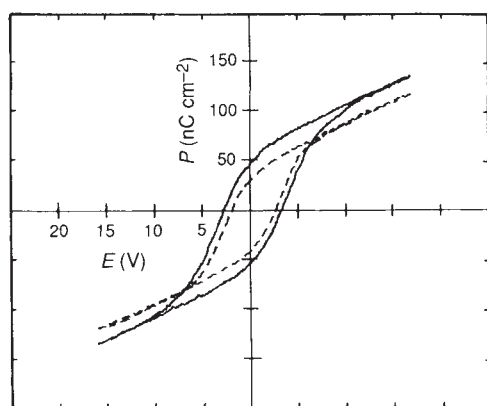
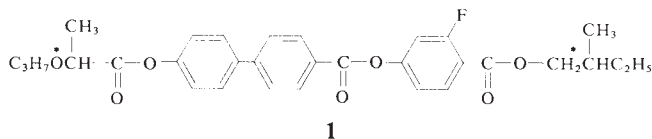


FIG. 2 Hysteresis loop of **P-E** curve for a mixture of FLC **1**/azobenzene **2**. Solid line, *trans* form of **2**; dotted line, *cis* form. The **P-E** curve was measured with a Sawyer-Tower capacitance bridge.

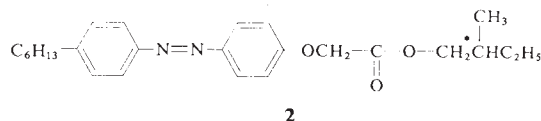
the polarization of each layer also spirals along the direction of the helical pitch and the overall polarization is averaged to zero. When the FLC molecules are placed in a cell with a small gap ($\sim 2 \mu\text{m}$), which is smaller than the helical pitch, the interaction of the FLC molecules and the cell surface is so strong that the helix structure becomes energetically unfavourable, and **n** in each layer is aligned in the same direction with the smectic layers oriented roughly perpendicular to the cell surface (Fig. 1a). As a result, the polarization of each layer is also aligned in one direction. This state is called the surface-stabilized state⁴. If an external electric field is applied across the cell, the polarization can be aligned upwards or downwards, depending on the polarity of the applied electric field (Fig. 1b). When the polarity of the electric field is reversed, the polarization of each FLC molecule flips (Fig. 1b). The two states are stable and remain unchanged even after the external electric field is removed. Therefore, the FLC in the surface-stabilized state exhibits bistability. The flip in polarization changes the director **n**, and hence the optical properties, as schematically illustrated in Fig. 1.

FLCs show much quicker response against the applied electric field owing to Ps, and exhibit a hysteresis between the applied electric field across the cell (**E**) and the polarization (**P**)^{2,3}. The shape of the hysteresis loop depends on the nature of each FLC^{2,3}.

We prepared a mixed phase incorporating the FLC **1** and the photosensitive azobenzene **2** (3 mol%), in a cell with a $2\text{-}\mu\text{m}$ gap, equipped with ITO electrodes.



Phase diagram, $\text{C}-38^\circ\text{C}-\text{SmC}^*-64^\circ\text{C}-\text{SmA}-1,10.5^\circ\text{C}-\text{I}$; spontaneous polarization, 82 nC cm^{-2} ; tilt angle, 20.1° .



Photoirradiation at 360 nm causes *trans-cis* isomerization of **2**. Before photoirradiation, when **2** was in the *trans* form, the mixture showed the hysteresis loop shown by the solid line in Fig. 2, whereas after photoirradiation when it was in the *cis* form, the mixture showed the hysteresis indicated by the dotted line in Fig. 2. The *trans* form is rod-like and tends to stabilize the LC phase, whereas the *cis* form is 'bent' and inclined to destabilize the LC phase¹⁶. Therefore, the *trans-cis* photoisomerization can cause disorganization of the phase structure

of the SmC^* phase, which may be the origin of the change in the coercive force of the mixture. This forms the basis for a novel photon-mode response of the FLC (see Fig. 3). (1) The polarizations of FLC were aligned in one direction by an external electric field ($\sim 10 \text{ V}$). (2) An electric field (**E**) was applied in the opposite direction to the original electric field, which was small enough to leave the initial polarization unchanged. (3) The cell was irradiated with light at $\sim 360 \text{ nm}$, which caused the *trans-cis* photoisomerization of the azobenzene. (4) This photochemical reaction lowered the coercive force of the mixture; when the coercive force fell below **E**, the polarization flipped into the opposite direction. The opposite direction of polarization produces different optical properties as described above; thus, an image can be stored in the mixed film.

The application of the original electric field and the *cis-trans* back isomerization which was accomplished by photoirradiation at $>450 \text{ nm}$ restored the initial direction of polarization, so the image can be easily erased and rewritten. Furthermore, as FLCs show bistability with the upward and the downward directions of polarization, an image stored in the form of contrast between these two states could be kept stable for a long period. A characteristic feature of the photochemical flip of polarization is the quick response. Time-resolved measurements with the aid of a pulsed laser (YAG laser, third harmonic; 10 ns full width at half maximum; 355 nm) have revealed that the polarization

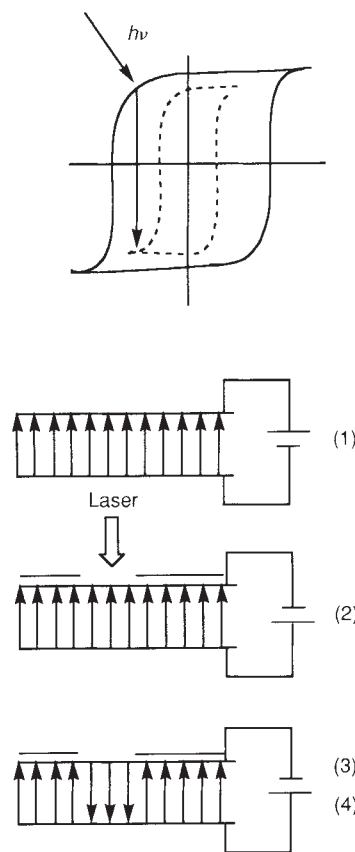


FIG. 3 Schematic illustration of photochemical flip of the polarization. The optical hysteresis loop changes on the photoisomerization of the azobenzene. (1) The polarization of FLC is aligned in one direction with the aid of the external electric field ($\sim 10 \text{ V}$). (2) An electric field (**E**) is applied in the opposite direction to the original electric field; the field is small enough to leave the initial polarization unchanged. (3) The cell is irradiated with light which caused the *trans-cis* photoisomerization of **2** ($\sim 360 \text{ nm}$). (4) The photochemical reaction of **2** simultaneously lowers the coercive force of the mixture, and when the coercive force drops below **E**, the polarization flips into the opposite direction. The different direction of **P** produces different optical properties; thus, an image can be stored in the mixture.

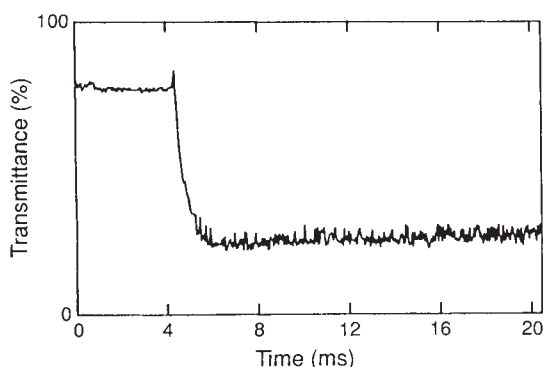


FIG. 4 Time-resolved measurement of the photochemical flip of polarization in the **2/1** mixture. Applied electric field: 4.8 V (opposite direction). Pulse source: YAG laser, 10 ns full width at half maximum at 355 nm. The change in transmittance of light from a He-Ne laser through a pair of crossed polarizers, with the sample cell between them, was measured with a photomultiplier as a function of time after a single pulse of the laser. The sample-cell temperature was maintained at 58 °C.

flip takes place on a timescale of several hundreds of microseconds (Fig. 4), which is two orders of magnitude faster than the photochemical phase transition. It is worth mentioning here that the *trans-cis* photoisomerization is a very fast process: we measured the rate of the photoisomerization of **2** in **1** and found that the photoisomerization was completed within a single pulse of the laser (<10 ns). It is therefore safe to say that the timescale of the photochemical reaction of the photosensitive molecule is infinitesimal in comparison with that of the flip of the polarization. The response depends on factors such as the position of the system of the hysteresis loop during photoirradiation, and

the structures of the photosensitive molecule and the FLC. The flip occurred more rapidly as the applied electric field in the opposite direction became larger, and azobenzenes with the same structure in chiral parts to the host FLC were more effective in inducing the flip. FLCs with a coercive force of ~0 V did not show a photochemical flip of polarization. A much faster response may be obtained by an appropriate choice of molecules.

A wide temperature range is available for the photochemical flip of polarization. The only requirement is that the LC shows the SmC* phase. The temperature range of the SmC* phase can be broadened by, for example, mixing several kinds of FLCs, so a wide range of working temperatures could be used for optical image storage systems based on this mechanism. □

Received 10 September 1992; accepted 5 January 1993.

1. Sasaki, A. *Molec. Cryst. Liq. Cryst.* **139**, 103–121 (1986).
2. Fukuda, A. & Takezoe, H. in *Structures and Properties of Ferroelectric Liquid Crystals* (Corona, Tokyo, 1990).
3. Skarp, K. & Handschy, M. A. *Molec. Cryst. Liq. Cryst.* **165**, 439–509 (1988).
4. Clark, N. A. & Lagerwall, S. T. *Appl. Phys. Lett.* **36**, 899–901 (1980).
5. Handschy, M. A. & Clark, N. A. *Ferroelectrics* **59**, 69–116 (1984).
6. Dewey, A. G. *Opt. Engng* **23**, 230–240 (1984).
7. Shibaev, V. P. et al. *Polym. Commun.* **24**, 364–365 (1983).
8. Coles, H. J. & Simon, R. *Polymer* **26**, 1801–1806 (1985).
9. Haas, W., Adams, J. & Wysocki, M. *Cryst. Liq. Cryst.* **7**, 371–379 (1969).
10. Haas, W. E., Nelson, K. F., Adams, J. E. & Dir, G. A. *J. electrochem. Soc.* **121**, 1667–1669 (1974).
11. Ogura, K., Hirabayashi, H., Uejima, A. & Nakamura, K. *Jpn. J. appl. Phys.* **21**, 969–973 (1982).
12. Odulov, S. G., Reznikov, Yu. A., Soskin, M. S. & Khizhnyak, A. I. *Sov. Phys. JETP* **58**, 1154–1158 (1983).
13. Gibbons, W. M., Shannon, P. J., Sun, S.-T. & Swetlin, B. J. *Nature* **351**, 49–50 (1991).
14. Kurihara, S., Ikeda, T., Sasaki, T., Kim, H.-B. & Tazuke, S. *J. chem. Soc. Chem. Commun.* 1751–1752 (1990).
15. Ikeda, T., Sasaki, T. & Kim, H.-B. *J. phys. Chem.* **95**, 509–511 (1991).
16. Ikeda, T., Horiuchi, S., Karanjit, D. B., Kurihara, S. & Tazuke, S. *Macromolecules* **23**, 42–48 (1990).

ACKNOWLEDGEMENTS. We thank K. Iimura and S. Ujije for discussions on the Sawyer–Tower bridge.

Large increases in flood magnitude in response to modest changes in climate

James C. Knox

234 Science Hall, Department of Geography, University of Wisconsin, Madison, Wisconsin 53706, USA

RECENT examinations of the possible hydrological response to global warming have emphasized changes in average conditions, rather than individual flooding events^{1–5}. Historical accounts suggest, however, that such events may have had a considerable regional impact^{6–9} even in the face of any relatively modest climate change⁸. Here I present a 7,000-year geological record of overbank floods for upper Mississippi river tributaries in mid-continent North America, which provides concrete evidence for a high sensitivity of flood occurrence to changing climate. During a warmer, drier period between about 3,300 and 5,000 years ago, the largest, extremely rare floods were relatively small—the size of floods that now occur about once every fifty years. After ~3,300 years ago, when the climate became cooler and wetter, an abrupt shift in flood behaviour occurred, with frequent floods of a size that now recurs only once every 500 years or more. Still larger floods occurred between about AD 1250 and 1450, during the transition from the medieval warm interval to the cooler Little Ice Age. All of these changes were apparently associated with changes in mean annual temperature of only about 1–2 °C and changes in mean annual precipitation of ≤10–20%.

Global-scale generalizations for relationships between extremes of floods and droughts and shifts in global atmospheric circulation often underestimate the true impact of climate

change because climate and hydrological responses vary regionally⁶. For example, many parts of Europe experienced anomalous wetness and great floods from about AD 1150 onward, and especially between AD 1300 and AD 1500, while southern Europe experienced very rare flooding⁷. A 1,000-year historical record of storms and large floods between AD 900 and 1900 in north China showed that large floods were infrequent during the comparatively warm Middle Ages (AD 1100–1250), but a modest climate change made them relatively frequent after AD 1250 into and during the comparatively cool climates of the Little Ice Age⁸. In Norway, however, the maximum incidence of increased flooding occurred during the Little Ice Age (AD 1650–1750)⁹.

The Upper Mississippi valley watersheds vary from a few to several hundred square kilometres and involve large overbank floods that mostly result from excess rainfall during summer, especially June and July when slow-moving cold fronts collide with moist and unstable tropical-derived air masses¹⁰. Other smaller overbank floods result from both snowmelt and more modest excess summer rainfall. The study region is climatically sensitive because it lies along a main seasonal climatic boundary that separates strongly contrasting air masses, and it includes an ecotone separating mixed hardwood forest to the northeast from prairie to the southwest^{11,12}. Holocene climate and vegetation changes have been documented in and near the study area, providing a framework from which hydrological responses to climate change can be assessed^{13–21}.

Cobble and boulder gravels transported by large overbank floods and now buried in overbank floodplain alluvium provided the basis for reconstructing a time series for large overbank floods. Former flood magnitudes were determined by computing the minimum flood depths ('competent' flood depths) capable of transporting the coarsest cobbles or boulders found in overbank flood deposits¹⁰.

Depths were calculated as

$$D = 0.0001 A^{1.21} S^{-0.57} \quad (1)$$

where D is competent flood depth in metres; A is intermediate axis (mm) represented by mean of coarsest five particles in a relict flood gravel; and S is approximate energy slope in metres per metre computed by estimating floodplain gradients for the channel reaches under investigation. The equation, which is statistically significant at the 0.0001 probability level with 74% explained variance, was calibrated from 52 observations of relationships between flow depth, particle size and slope^{22–26}.

Figure 1 shows the reconstructed record of long-term variations in late Holocene flood magnitudes, represented by relict flood water depths (d) expressed as a ratio to the corresponding bankfull stage water depth (d_{bkt}) at the same site of the sampled flood event. The bankfull depth is defined as the difference between the top of a channel-bottom lag gravel and the top of lateral accretion point bar sediment at the same site. Statistical regression with modern data indicates that the standard error for bankfull depth definition is in the range 15–18%. Stratigraphic excavations accompanied by radiocarbon dating showed that the channel-bottom lag gravel has maintained a uniform cross-valley elevation throughout the Holocene and that channel-bottom cutting and filling have been insignificant. The bed level at any given site thereby provided a reference against which relict flood depths could be compared. Flood depths at each given site were compared to corresponding site bankfull depths and expressed as dimensionless ratios to provide a composite presentation of regional flood events (Fig. 1).

I used 114 field sites representing boulder or cobble gravels deposited by ~68 Holocene overbank floods in southwestern Wisconsin to construct the flood record of Fig. 1. Most field sites involved excavation of buried gravels exposed in stream cutbanks. The age assignment for relict flood gravels was determined either directly or indirectly from radiocarbon dating. Direct age determinations from organic matter contained within flood deposits were available at only three sites because sufficient organic carbon suitable for radiocarbon dating was seldom found in the oxidizing environments of the overbank sediments. On the other hand, ages for nearly all of the stratigraphic units underlying the flood gravels were defined by radiocarbon dates. Most flood gravel ages were determined from a regression equation relating thickness of overbank Holocene floodplain sediment deposition and drainage area to time in radiocarbon years for 34 dated floodplain sites in the study area²⁷. The equation accounted for ~89% of the variance in the base data,

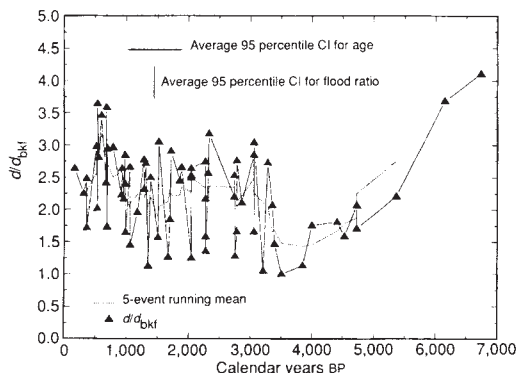


FIG. 1 Geological record of Holocene overbank floods in southwestern Wisconsin. The average magnitude for the interval separating the upper and lower 95th percentile confidence intervals (CI) is ~0.6 for the flood ratio and 870 years for the age estimate. A flood ratio of 2.0 corresponds to modern floods expected about once every 30–50 years whereas a flood ratio of 3.0 relates to modern floods expected about once every 500–1,000 years in southwestern Wisconsin.

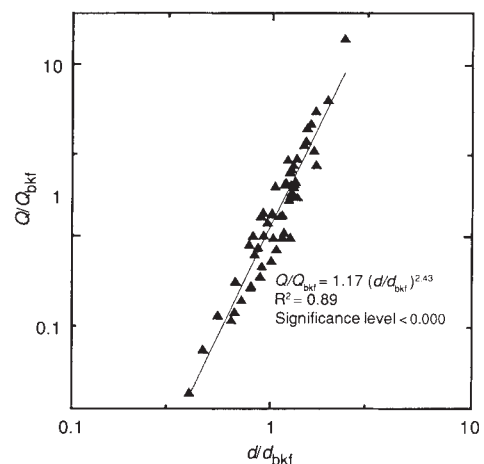


FIG. 2 Example dimensionless rating curve. Results are for Platte River near Rockville, Wisconsin; U.S. Geological Survey gage no. 0541400. Discharge of a flood Q , is expressed as a ratio to the bankfull discharge Q_{bkt} , and depth of a flood is expressed as a ratio to the bankfull depth. The coefficient and exponent in the equation are close to the values of 1.0 and 2.5 expected from channel hydraulic geometry theory, whereby the modal value for the power function rate of increase of depth with discharge in a channel cross-section is 0.4.

and the mean difference between the upper and lower 95th percentile confidence intervals was 870 years. Radiocarbon ages were subsequently converted to calendar year ages using the method of Stuiver and Reimer²⁸.

A striking illustration of non-stationarity for the mean and variance of overbank floods is indicated by the long-term change in flood magnitudes for the events (Fig. 1). Maximum values for the ratio of recorded overbank floods to their corresponding bankfull stage floods have been much larger for the past 3,300 years than they were ~3,300–5,000 years ago. During the past 3,300 years, overbank floods have frequently approached, and in a few instances exceeded, depths that were three times the magnitude of their associated bankfull stages (Fig. 1). Between 3,300 and 5,000 years ago overbank floods were typically less than twice the depth of the corresponding bankfull stage (Fig. 1). The significance of the apparent two large flood events between 6,000 and 7,000 years ago is unclear because of the small sample size, although an independent study in this region indicated that high-frequency bankfull stage floods were also anomalously large during that interval²⁷. Figure 1 data also suggest that flood magnitudes increased after ~1,000 years ago, and especially ~700–500 years ago during the transition from the medieval warm interval into the early Little Ice Age millennialia. The data in Fig. 1 suggest that even the smaller, more-frequent overbank floods were larger after ~1,000 years ago than during earlier Late Holocene times.

Because of the stable channel base elevation for study area streams, the expected exceedence probabilities of relict flood events shown on Fig. 1 can be estimated from a dimensionless rating curve relating discharge to flood depth. Figure 2, representing a site in the central study area, illustrates how a dimensionless flood ratio d/d_{bkt} can be used to predict flood discharge. Conventional log-Pearson III methodology for flood frequency determination indicated that expected recurrence intervals for discharge ratios on Fig. 2 are associated with flood magnitudes that range from more than once per year to roughly once in the 500 years^{29,30}. The analysis indicates that flood depth ratios with a magnitude of ~2.0 correspond with a modern flood being expected once per 30–50 years, whereas flood depth ratios of 2.5 and 3.0 relate to modern floods expected once in ~150–200 years and once in ~500–1,000 years, respectively.

Translating flood depth ratios into expected recurrence frequencies, and comparing flood magnitudes and frequencies shown on Fig. 1 with regional fossil proxy measures of climate, indicates that modest climate changes have strongly influenced the recurrence probabilities of large floods. For example, the period of small floods between 5,000 and 3,300 years ago, when the largest floods could be equated with modern floods of 2% recurrence probability, corresponds to a dry climate, with proxy data suggesting precipitation ~15% less than modern values and a July temperature at least 0.5 °C warmer than before or after¹⁹. The occurrence of more-frequent large floods after ~3,300 years ago also agrees with proxy climate data indicating that annual precipitation may have increased by ~8% and mean July temperature decreased by ~0.7 °C between ~4,000 and 3,000 years ago, and that between ~2,000 and 1,500 years ago annual precipitation increased by an additional 8% to roughly its modern magnitude, whereas mean July temperature increased gradually by ~0.4 °C (ref. 13). In the Upper Mississippi valley

more-frequent recurrences of large floods are particularly noted after 1,000 years ago and especially between ~700 and 500 years ago. The late-Holocene pattern of magnitude variability for bankfull stage floods of 1–2 years recurrence frequency in this region²⁷ was similar to that reported here for the larger overbank floods.

I conclude that modest climate changes, generally smaller than climate changes predicted by global circulation models for greenhouse gas increases, caused large and sometimes abrupt adjustments in both magnitudes and frequencies of floods in the Upper Mississippi valley. This sensitivity casts doubt on the assumption of homogeneity of the mean and variance, used in many standard statistical techniques for estimating long-term potential flood magnitudes and frequencies. It also suggests that the magnitudes and frequencies of floods in many regions could change considerably if current trends in emissions of atmospheric greenhouse gases continue. □

Received 21 September; accepted 21 December 1992.

- Rind, D. *et al.* *Nature* **358**, 119–122 (1992).
- Mimikou, M. A. *et al.* in *Hydrology for the Water Management of Large River Basins* (eds Van de Ven, F. H. M. *et al.*) 173–182 (Int. Assoc. Hydrol. Sci. Publ. No. 201, 1991).
- Rind, D. *J. geophys. Res.* **93**, 5385–5412 (1988).
- Gleick, P. H. *Rev. Geophys.* **27**, 329–344 (1989).
- Lettenmaier, D. P. & Gan, T. Y. *Water Resources Res.* **26**, 69–86 (1990).
- Folland, C. K. *et al.* in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T. *et al.*) 195–238 (Cambridge Univ. Press, 1990).
- Lamb, H. H. *Climate: Present, Past and Future—Vol. 2, Climatic History and the Future* 143–145 (Methuen, London, 1977).
- Zhang, D. in *Quaternary Geology and Environment in China* (ed. Liu, T.) 208–213 (Science Press, Beijing, 1991).
- Grove, J. M. *Arct. Alp. Res.* **4**, 131–138 (1972).
- Knox, J. C. in *Flood Geomorphology* (eds Baker, V. R. *et al.*) 279–300 (Wiley, New York, 1988).
- Bryson, R. A. *Geogr. Bull.* **8**, 228–269 (1966).
- Bernabo, J. C. & Webb, III, T. *Quat. Res.* **8**, 64–96 (1977).
- Bartlein, P. J. *et al.* *Quat. Res.* **22**, 361–374 (1984).
- Chumbley, C. A. *et al.* *Science* **249**, 272–274 (1990).
- Davis, A. M. *Ann. Assoc. Am. Geog.* **67**, 204–213 (1977).
- Keen, K. L. & Shane, L. C. K. *Geol. Soc. Am. Bull.* **102**, 1646–1657 (1990).

- King, J. E. *Ecol. Monogr.* **51**, 43–62 (1981).
- Maher, L. J. in *Quaternary History of the Driftless Area* (eds Knox, J. C. *et al.*) 119–135 (Wisconsin Geology & Natural History Survey, 1982).
- Winkler, M. G. *et al.* *Quat. Res.* **25**, 235–250 (1986).
- Wright, Jr. H. E. *et al.* *Geol. Soc. Am. Bull.* **74**, 1371–1396 (1963).
- Baker, R. G. *et al.* *Quat. Res.* **37**, 379–389 (1992).
- Lane, E. W. & Carlson, E. J. *Proc. 5th Int. Assoc. Hydraul. Res. Congr.* 37–48 (1953).
- Wolman, M. G. & Eiler, J. P. *Trans. Am. geophys. Un.* **39**, 1–14 (1958).
- Fahnestock, R. K. *U.S. Geol. Surv. Prof. Pap.* 422-A (U.S. Government Printing Office, 1963).
- Kellerhals, R. J. *Waterways, Harbors Coastal Eng. Div., Am. Soc. Civ. Eng.* **93**, 63–84 (1967).
- Scott, K. M. & Gravlee, G. C. Jr. *U.S. Geol. Surv. Prof. Pap.* 422-M (U.S. Government Printing Office, 1968).
- Knox, J. C. *Quat. Res.* **23**, 287–300 (1985).
- Stuiver, M. & Reimer, P. J. *Radiocarbon* **28**, 1022–1030 (1986).
- Guidelines for Determining Flood Flow Frequency Bull. 17B* (U.S. Water Resources Council, Washington DC, 1981).
- Dunne, T. & Leopold, L. B. *Water in Environmental Planning* 646–649 (Freeman, San Francisco, 1978).

ACKNOWLEDGEMENTS. This research was supported by the NSF. I thank F. J. Magilligan and R. D. Dunning for field assistance.

The 'flickering switch' of late Pleistocene climate change

K. C. Taylor*, G. W. Lamorey†, G. A. Doyle‡, R. B. Alley‡, P. M. Grootes§, P. A. Mayewski||, J. W. C. White¶ & L. K. Barlow¶

* Desert Research Institute, University of Nevada System, Reno, Nevada 89506, USA

† Hydrology/Hydrogeology Science Program, University of Nevada, Reno, Nevada 89557, USA

‡ Earth System Science Center and Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

§ Department of Geological Sciences and Quaternary Research Center, University of Washington, Seattle, Washington 98195, USA

|| Glacier Research Group, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire 03824, USA

¶ Institute of Arctic and Alpine Research and Department of Geological Sciences, University of Colorado, Boulder, Colorado 80309, USA

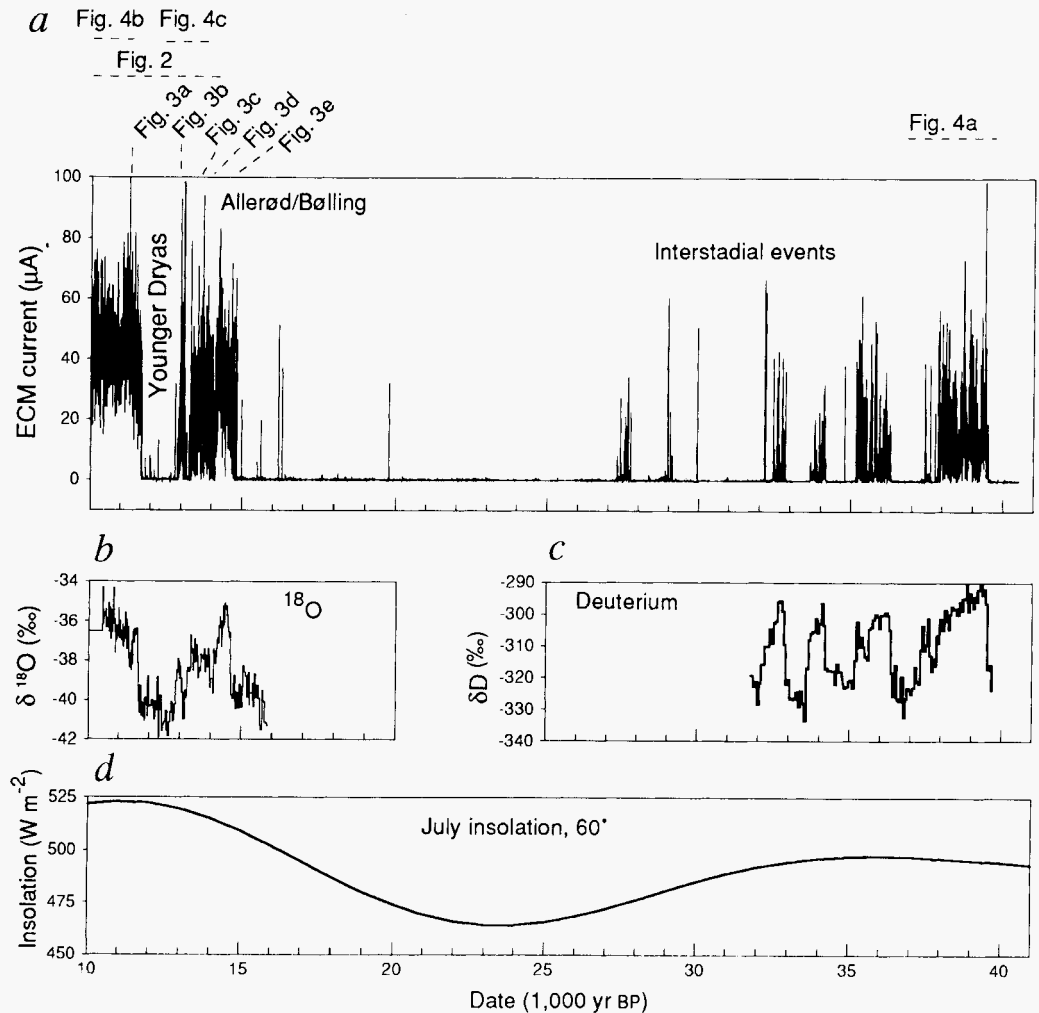
POLAR ice contains a unique record of past climate variations; previous Greenland ice cores have documented relatively warm 'interstadial' periods during the last glaciation and short (century-scale) returns to colder conditions during the glacial to interglacial warming (see, for example, ref. 1). These climate features have also been observed to varying degrees in ocean sediment cores^{2–4} and terrestrial pollen and insect records^{5–7}. Here we report electrical conductivity measurements from a new Greenland ice core, which confirm these previous observations, and also reveal a hitherto unrecognized mode of rapid climate variation. Fluctu-

ations in ice conductivity on the scales of <5–20 years reflect rapid oscillations in the dust content of the atmosphere. This 'flickering' between two preferred states would seem to require extremely rapid reorganizations in atmospheric circulation.

The Greenland Ice-Sheet Project 2 (GISP2) is coring the summit of the Greenland ice sheet (72.6° N, 38.5° W) to develop a continuous high-resolution record of climate related parameters going back at least 200,000 years. The electrical conductivity measurement (ECM) measures the ability of an ice core to conduct an electrical current, which is related to the balance of acids and bases in the ice^{8–10}. The measurement is made by determining the current flowing between two moving electrodes with a potential difference of a few thousand volts. The current increases with increased concentration of strong acids, especially sulphuric (from volcanic activity⁸ and other sources¹¹) and nitric (controlled mainly by atmospheric chemistry^{12,13}). The current decreases when the acids are neutralized, which is most commonly caused by ammonia (biomass burning and other sources^{14,15}) or due to alkaline dust (continental sources¹²).

Changes in the value of the ECM baseline occur when climate or weather patterns change the source strength or transport of the acids and bases. During Wisconsinian times, dust transport to Greenland increased by at least a factor of 40, possibly because of changes in transport paths, wind speeds and surface moisture¹⁶. The large amounts of alkaline dust present in ice from cold periods during the last glacial period neutralize the acids, reducing the current flow between the electrodes by a factor of ~40 compared with Holocene conditions. This permits clear discrimination between dusty and less dusty conditions. The relationship between the ECM signal and the concentration of alkaline dust at the transition from acidic to alkaline conditions (0 to 10 µA) is not well quantified. At the transition between

FIG. 1 *a*, GISP2 electrical conductivity measurement (ECM) record for the period of 10–40.5 kyr BP. Low current levels indicate that alkaline dust has neutralized the acidity of the ice. The record is resampled to a rate of one sample per year. The ages of the other expanded-scale figures (Figs 2–4) are indicated. The years indicated are calendar years. We adopt the carbon-14 datum of AD 1950 as present time. *b*, Isotopic ratio $^{18}\text{O}/^{16}\text{O}$ for the ages 10.5 to 16 kyr BP. *c*, Isotopic ratio D/H for the ages 32 to 40.5 kyr BP. The isotope samples are from contiguous 1-m sections of the core. Because of flow-induced thinning of annual layers, the 1-m sections correspond to one sample every 15 to 37 years in the $^{18}\text{O}/^{16}\text{O}$ record and one sample every 65 to 80 years in the deuterium/hydrogen record. The correlation between the ECM and the isotope records demonstrates that the ECM record is responding to climatic events. *d*, Mean July insolation for latitude 60° north²³.



acidic and alkaline conditions, the effect of small amounts of additional dust on the magnitude (but not the frequency) of the ECM signal may be disproportionately large. In the GISP2 core, however, detailed chemical analysis shows that the order-of-magnitude decreases in ECM that we consider here are associated with order-of-magnitude increases in calcium. The source for the calcium is believed to be airborne calcium carbonate

dust. The ECM record has the highest time resolution (>15 samples per year) of available measurements, and together with the sensitivity to dust, this makes ECM suitable for investigating the rate of rapid atmospheric circulation changes.

The ECM record discussed here covers the time period 10 to 42 thousand years before present (kyr BP), and allows climate variability to be investigated on timescales of seasons to

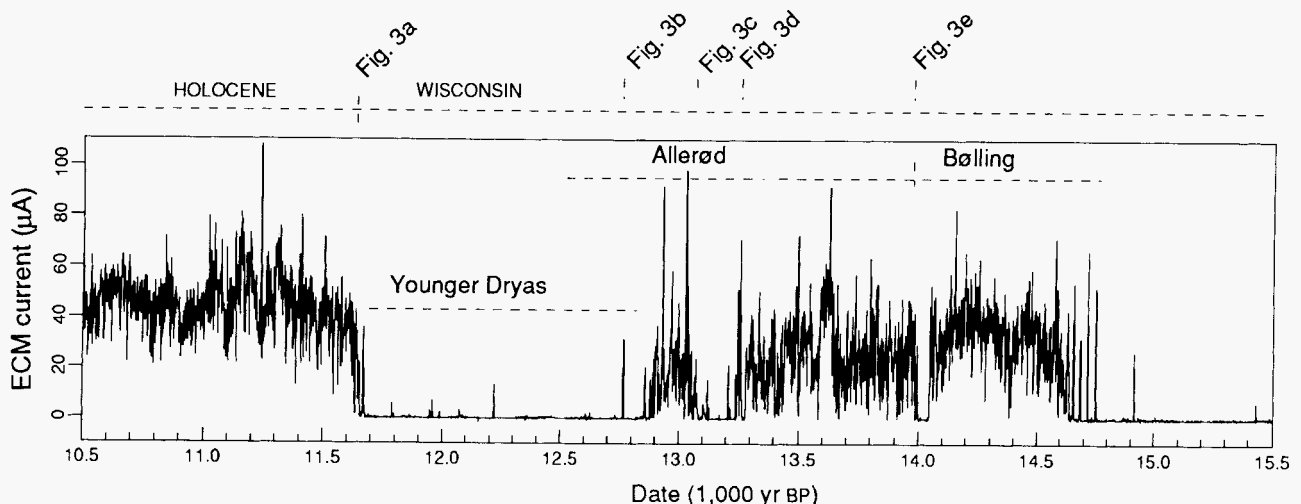


FIG. 2 Yearly averages of ECM in the GISP2 core at Summit, Greenland for the period of 10.5 to 15.5 kyr BP. The ages of expanded figures

(Fig. 3a–e) are indicated.

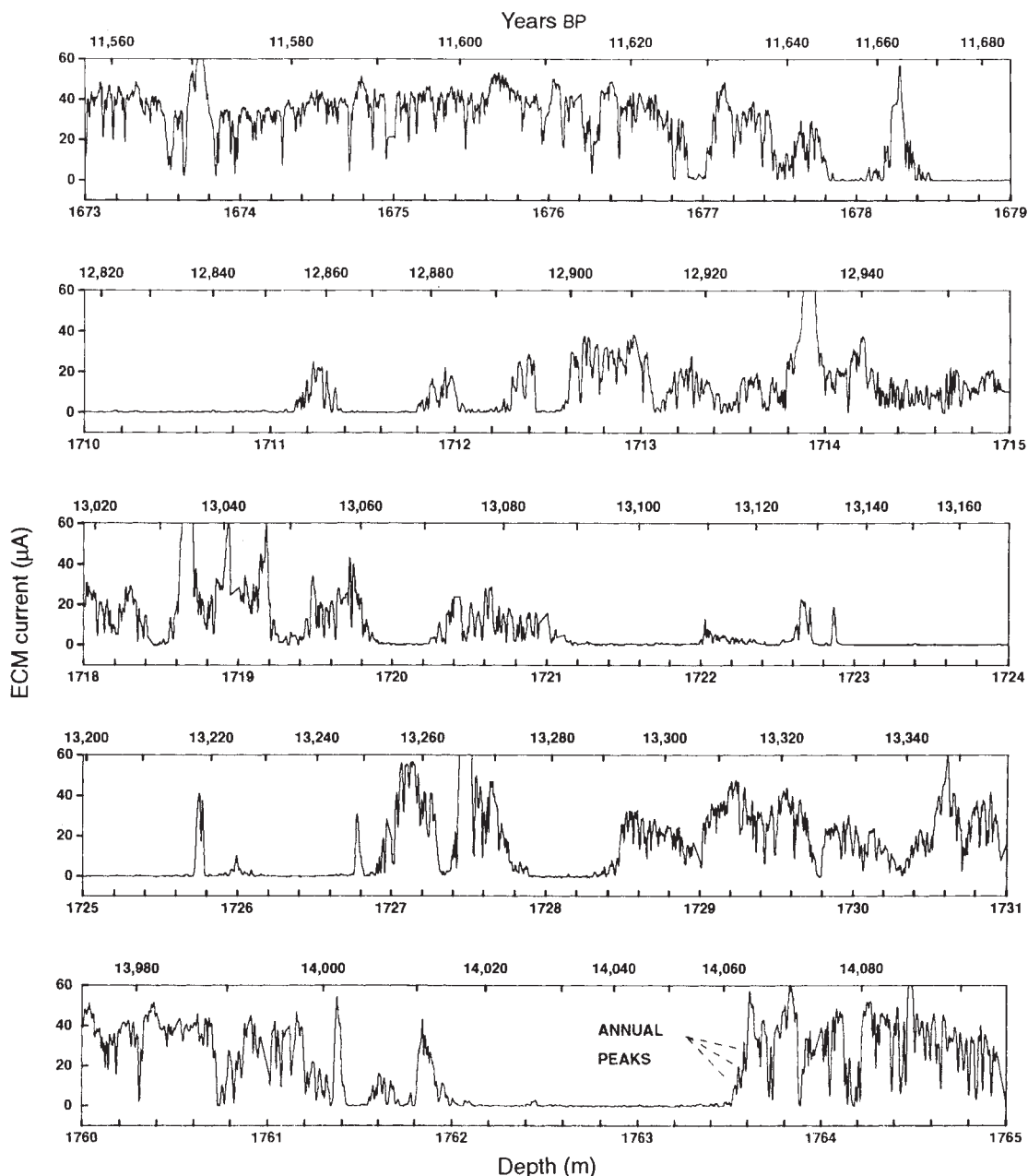
millennia (Fig. 1a). The core has been preliminarily dated by counting of annual layers which are identified by visual stratigraphy, ECM and particulate measurements. Every annual layer has been counted for the past 17.4 kyr BP with an estimated accuracy of 3% (ref. 17). Between 17.4 and 40.5 kyr BP, annual layers were counted by visual stratigraphy for one out of every 10 or 20 m. The interim sections were dated by counting interpolated annual layers. At an age of 15 kyr BP, ages of identical events in the GISP2 core and GRIP core¹ (located 30 km to the east) differ by 200 years. This is within expected errors. At an age of 40 kyr BP in the GISP2 core, corresponding events in the GRIP core are dated at 34 kyr BP. Both groups are refining the preliminary dating of the cores and this discrepancy is likely to diminish. A comparison of the two cores will be presented elsewhere.

The ECM record is characterized by a bimodal high or low pattern that is controlled by the absence or presence of alkaline dust. For comparison we also present stable isotope profiles, which are known to respond to the temperature difference between the moisture source and the ice deposition site, and are used extensively for palaeoclimate reconstructions^{18–20}. The

synchronous transitions in the ECM and isotope records (Fig. 1b, c) demonstrate the sensitivity of the ECM to climate conditions. Periods of heavy isotopic composition indicate warmer interglacial or interstadial conditions. The high ECM values during these times indicate that it was also less dusty. The ECM record (Fig. 2) clearly shows well known millennium-scale Younger Dryas, Allerød, Bølling and Interstadial periods, and century-scale events during the Allerød and Bølling periods. The isolated spikes on the ECM record span at most a few years and are likely to be associated with volcanic activity. Although isotopes and ECM show synchronous transitions in Fig. 1, they do not correlate well on timescales of a few years. This indicates that they respond to different aspects of climate (temperature and dust) rather than there being a direct influence of atmospheric particulates on the isotopic composition of snow fall^{21,22}. Ice rheology has been observed to covary with ECM in the GISP2 core and other²³ Arctic cores. This phenomenon is most probably due to effects of particulates on grain recrystallization²⁴, not to an influence of ice rheology on ECM.

In addition to the well known millennium-scale stadial and interstadial periods (best viewed in Fig. 1), and the previously

FIG. 3 a–e, Expanded ECM records of recent climate transitions plotted against depth (bottom axis) with one sample per millimetre of depth. a, Start of Holocene, end of Younger Dryas. b, Start of Younger Dryas. c, End of inter-Allerød cold period. d, Start of inter-Allerød cold period. e, Transition. The top axes are age, in years BP. At this scale the annual signal in the ECM can occasionally be resolved as small local peaks that occur once per year. During the transition at 14,060 BP (e), three annual peaks can be resolved. The annual layer thickness is ~6 cm per year during warmer, less dusty conditions and 3 cm per year during colder dusty conditions. Occasionally the number of ECM annual peaks and annual layers in a given segment differ slightly because particulate and visual stratigraphy records were also considered when annual layers were identified.



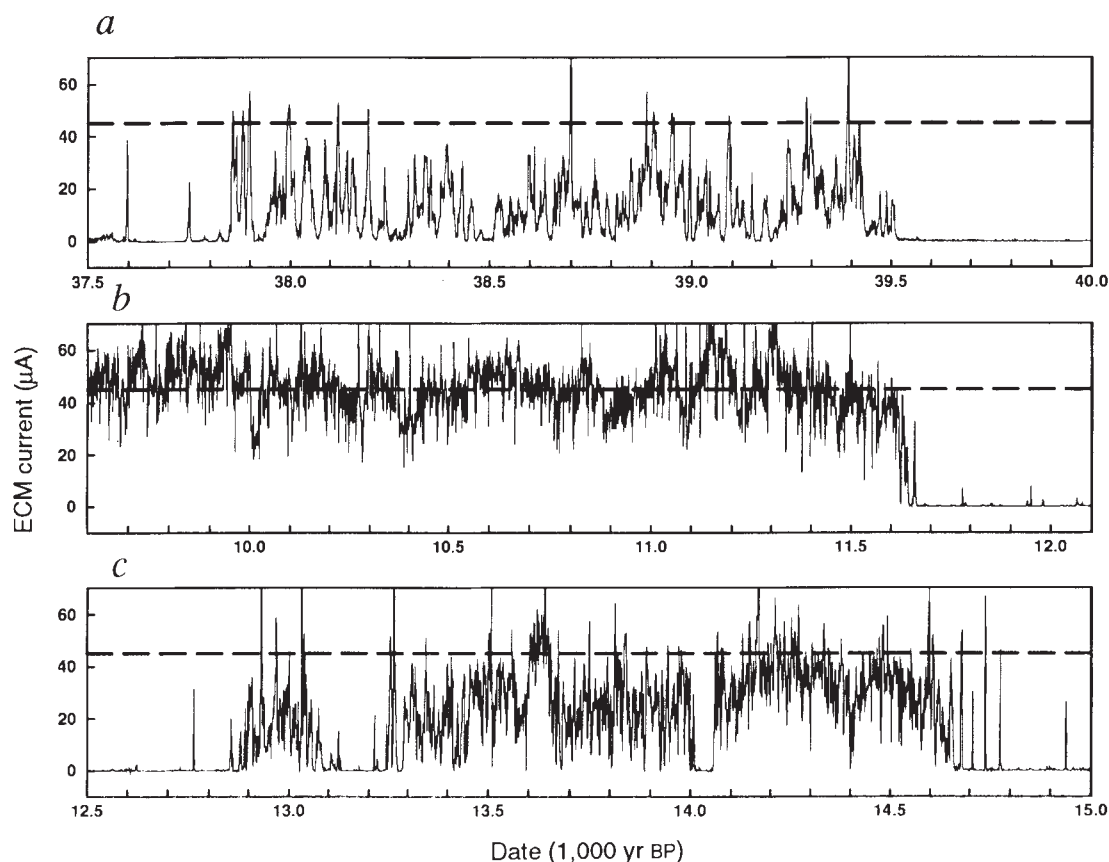


FIG. 4 Yearly averaged ECM profiles for three 2,500-year intervals. *a*, Interstadial event at 37.8 to 39.5 kyr BP; *b*, the start of the Holocene; *c*, the Allerød and Bølling periods.

recognized century-scale climate events that occur during the Allerød and Bølling periods (best viewed in Fig. 2), we detect a still higher frequency of variability associated with abrupt climate change (Fig. 3*a–e*). The transitions between century or longer warm and cold periods are characterized by abrupt fluctuations in alkaline dust concentrations that occur over periods of less than 10 years. There are frequent brief (10–20 years) and abruptly terminating (less than 5 years) periods during which the ECM returns to pre-transition values. Although a similar timescale has been reported for temperature change at the end of the Younger Dryas (ref. 25), the fluctuations we have observed are quite different from the smooth transitions previously reported. During the Interstadial periods the fluctuations are continuous (Fig. 4*a*), in sharp contrast to the more stable Holocene (Fig. 4*b*). The Allerød and Bølling periods (Fig. 4*c*) are more stable than the Interstadial periods but not as stable as the Holocene. The transitions at the start and end of the interstadial periods occur within a decade. During the interstadial events and at main climate transitions, the part of the climate system influencing the ECM is behaving like a flickering switch that fluctuates between two states before stabilizing. The instability of the Interstadial events may be related to a prolonged intermediate value of July insolation at 60° north latitude²⁶ (Fig. 1*d*), which contrasts with the more stable cold climate that occurs during the low-insolation period. During the Holocene, when climate and the ECM are relatively stable, insolation decreases from 523 W m⁻² to 475 W m⁻² (ref. 26), indicating that insolation alone does not control climate stability.

The seasonal time resolution of the ECM record portrays an aspect of the climate system that consistently and frequently changes between glacial and near-interglacial conditions in periods of less than a decade, and on occasion as quickly as three years. Instead of being rapid and smooth, the transitions are characterized by flickering between preferred states. It is improbable that the surface area of dust sources such as loess-covered land, freshly exposed by retreating ice or water, can change as abruptly as is observed in the ECM record. A re-ordering of

atmospheric circulation could rapidly change the surface moisture, wind speeds and air-mass routing and, hence, change the quantity of airborne loess. A re-ordering of atmospheric circulation and increased wind speeds is consistent with observed changes in the temperature gradient between high and mid northern latitudes during glacial conditions²⁷ and with evidence from ocean sediment cores showing rapid changes in sea surface temperatures³. The influence of temperature gradients on atmospheric circulation suggests that the rapid changes observed in the ECM record are a result of changes in the mechanisms influencing high- to mid-latitude Northern Hemisphere temperature gradients. Concepts of mechanisms for rapid changes in ocean and atmospheric circulation, and in sea ice extent, need to be examined to see if they can explain the repeated and rapid changes observed in this record. As additional records from the new Greenland ice cores become available it will be possible to develop and test hypotheses regarding climate mechanisms responsible for the abrupt and frequent climate changes we observe. The increased time resolution provided by these cores may allow analogies to be drawn with the instrumental record, thus opening other ways to understand the mechanism responsible for the flickering switch. □

Received 5 October; accepted 22 December 1992.

- Johnsen, S. *et al.* *Nature* **359**, 311–313 (1992).
- Lehman, S. J. & Keigwin, L. D. *Nature* **356**, 757–762 (1992).
- Ruddiman, W. F. & Macintyre, A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **35**, 45–214 (1981).
- Broecker, W. S. *et al.* *Palaeoceanography* **5**, 469–477 (1990).
- Wolillard, G. M. & Mode, W. D. *Science* **215**, 159–161 (1982).
- Kolstrup, E. *Geologie Mijnb.* **69**, 253–262 (1990).
- Coope, G. R. *Phil. Trans. R. Soc. B*, **280**, 313–2440 (1977).
- Hammer, C. *et al.* *J. Glaciol.* **25**, 359–372 (1980).
- Taylor, K. C. *et al.* *J. Glaciol.* **38**, 325–332 (1992).
- Moore, J. C. *et al.* *J. geophys. Res.* **97**, 1887–1896 (1992).
- Bates, T. S. *et al.* *J. Atmos. Chem.* **14**, 315–337 (1992).
- Neftel, A. *et al.* *Am. geophys. Un. Geophys. Monogr.* **33**, 32–38 (1985).
- Legrand, M. & Kirchner, S. *J. geophys. Res.* **95**, 3493–3509 (1990).
- Legrand, M. *et al.* *Geophys. Res. Lett.* **19**, 473–477 (1992).
- Logan, J. *et al.* *J. geophys. Res.* **88**, 10785–10807 (1983).
- Hammer, C. *et al.* *Am. geophys. Res. Geophys. Monogr.* **33**, 90–95 (1985).
- Meese, D. A. *et al.* *U.S. Army Cold Regions Res. Engng Lab. Rep.* (in the press).

18. Dansgaard, W. *Tellus* **16**, 436–467 (1964).
19. Lorius, C. *et al. Nature* **280**, 644–648 (1979).
20. Dansgaard, W. *et al. Science* **218**, 1273–1277 (1982).
21. Fischer, D. A. *Tellus* **B43**, 401–407 (1991).
22. Fischer, D. A. *The Last Deglaciation: Absolute and Radiocarbon Chronologies*, NATO ASI Series, Vol. 12, 200–212 (1992).
23. Fischer, D. A. *The Physical Basis of Ice-sheet Modelling*, IAHS Publ. 170, 45–51 (1987).
24. Paterson, W. S. B. *Cold Regions Sci. Technol.* **20**, 75–98 (1990).
25. Dansgaard, W. *et al. Nature* **339**, 532–523 (1989).
26. Berger, A. & Loutre, M. *Quat. Sci. Rev.* **10**, 297–317 (1991).
27. Frenzel, B. *et al. (eds) Atlas of Palaeoclimates and Palaeoenvironments of the Northern Hemisphere*, 39–45 (Geographical Research Institute, Budapest, 1992).

ACKNOWLEDGEMENTS. We thank the NSF Division of Polar Programs for funding, numerous GISP colleagues, the Science Management Office (University of New Hampshire), the Polar Ice Coring Office (University of Alaska) and TAG 109th (New York Air National Guard) for their assistance.

Reduction of Fe(III) in sediments by sulphate-reducing bacteria

Max L. Coleman^{*†}, David B. Hedrick[‡],
Derek R. Lovley[§], David C. White^{‡||} & Kenneth Pye^{*}

^{*} Postgraduate Research Institute for Sedimentology,
University of Reading, Whiteknights, PO Box 227, Reading RG6 2AB, UK
[†] BP Research & Engineering Centre, Sunbury-on-Thames,
Middlesex TW16 7LN, UK

[‡] Center for Environmental Biotechnology, University of Tennessee,
10515 Research Drive, Knoxville, Tennessee 37932-2567, USA

[§] Water Resources Division, 430 National Center,
United States Geological Survey, Reston, Virginia 22092, USA

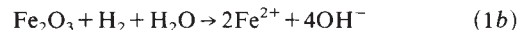
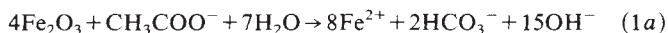
^{||} Microbiology Department, University of Tennessee, Knoxville,
Tennessee 37996-0845 and Environmental Science Division,
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

REDUCTION of ferric iron (Fe(III)) to ferrous iron (Fe(II)) is one of the most important geochemical reactions in anaerobic aquatic sediments because of its many consequences for the organic and inorganic chemistry of these environments¹. In marine environments, sulphate-reducing bacteria produce H₂S, which can reduce iron oxyhydroxides² to form iron sulphides. The presence of siderite (FeCO₃) in marine sediments is anomalous, however, as it is unstable in the presence of H₂S. Previous work^{3,4} has suggested a bacterial origin of siderite. Here we describe geochemical and microbiological studies which suggest that contemporary formation of siderite concretions in a salt-marsh sediment results from the activity of sulphate-reducing bacteria. We find that, instead of reducing Fe(III) indirectly through the production of sulphide, some of these bacteria can reduce Fe(III) directly through an enzymatic mechanism, producing siderite rather than iron sulphides. Sulphate-reducing bacteria may thus be an important and previously unrecognized agent for Fe(III) reduction in aquatic sediments and ground waters.

Nodular concretions of intimately mixed siderite and FeS are growing within sulphide-rich, Recent, salt-marsh sediments on the coast of Norfolk, UK⁵. The host mud around the concretions contain FeS but little siderite. Although modelled growth rates⁶ for carbonate concretions imply that formation may take > 10⁵ years, these concretions are found in sediments that have been deposited in the last 50 years. This makes them uniquely suitable for investigations into the actual mechanisms of siderite concretion formation.

The formation of siderite in organic-rich marine mudrocks has previously been associated with the degradation of organic matter by anaerobic, methanogenic bacteria⁷. Carbonate resulting from this process has a characteristic positive isotope signature ($\delta^{13}\text{C} = +10$ to $+15\%$ PDB)⁸. However, the siderite concretions in the marsh have $\delta^{13}\text{C}$ values averaging -5.9% PDB, suggesting that the siderite is not formed under methanogenic conditions. An alternative microbial process that could favour siderite formation is microbial reduction of Fe(III). Most of the microbial Fe(III) reduction in aquatic sediments is proposed to

result from the oxidation of important fermentation products such as acetate and hydrogen¹. (The equations below illustrate the overall mass balance and may represent many contributory reactions).



The production of ferrous iron, hydroxyl ions and bicarbonate should all favour siderite formation



To learn whether microbial Fe(III) reduction played a role in siderite formation, we analysed the microbial populations in a concretion and in two host-sediment samples from the marsh. The concretion was subspherical, ~7 cm in diameter and dark grey to black, its host an anoxic, black clayey-silt. As we cannot yet do all analyses on one sample, we have assumed that the mineralogy was similar to that of visually identical samples, which were indistinguishable from those reported previously from the same location (finely intergrown siderite, Mg-calcite and iron sulphide⁵).

Lipid biomarker analyses provides *in situ* measures of sedimentary microbial community structure and nutritional status without the requirement for isolation and growth of the microbiota⁹. Polar lipid fatty acids (PLFA) are maintained metabolically¹⁰ and are an accurate measure of viable sedimentary microbial biomass¹¹. Viable eubacterial biomass in the host-mud, measured as the PLFA per g dry weight of sediment, was

TABLE 1 Microbial biomarkers in the concretion and host mud

Parameter	Concretion	Host mud	Interpretation for the concretion, relative to host mud
PLFA (μg per g dry wt)	0.7 1.4–7.0*	4.8, 5.4	Two-thirds of the eubacterial biomass
PLDE (ng per g dry wt)	N.D.†	10, 12	No detectable methanogens
MK/PLFA (g per g)	0.27	0.06, 0.14	Much larger proportion of anaerobic respirers, relative to fermenters
UK/PLFA (g per g)	N.D.†	N.D.†	No detectable aerobes, in either
TL/PLFA (g per mg)	1.0	0.16, 0.22	More accumulated total lipid (former activity)
10-Methyl-16:0 (mol %)	2.0	2.4, 2.4	Less <i>Desulfobacter</i>
i17:1 (mol %)	1.5	0.8, 1.0	More <i>Desulfovibrio</i>

* Corrected for lower porosity.

† Not detected.

The concretion was packed in host anaerobic mud for transport to the laboratory. Despite this precaution it was found to have developed a brownish-orange rim as a result of atmospheric exposure. It was fractured with a solvent-cleaned chisel, and the outer orange layer discarded. The hard core of the concretion was pulverized with a mortar and pestle, and sieved through a 1-mm screen. Two samples of the host mud and the core were lyophilized. Lipid was extracted with a one-phase chloroform:methanol extractant, and fractionated into neutral lipid, glycolipid and polar lipid fractions with silicic acid. After derivatization, the eubacterial PLFA were analysed by capillary gas-chromatography/mass spectrometry (GC/MS)^{9,13}, and the archaeobacterial PLDE and PLTE by supercritical fluid chromatography¹⁴. Ubiquinones and menaquinones were recovered from the neutral lipid fraction of the lipid extracts, separated on thin layer plates, recovered, and analysed by supercritical fluid chromatography. Bacteria grow in the water in pore space, and precipitation of concretionary cement in pore space in the mud reduces the volume of fluid available. Mineral and chemical analyses⁵ suggest that porosity in the mud is 2–10 times greater than in the concretion, so we have multiplied the PLFA values measured in the latter by those factors (all other measures are relative) to allow comparison of microbial activities. Details of species identification are given in Fig. 1.

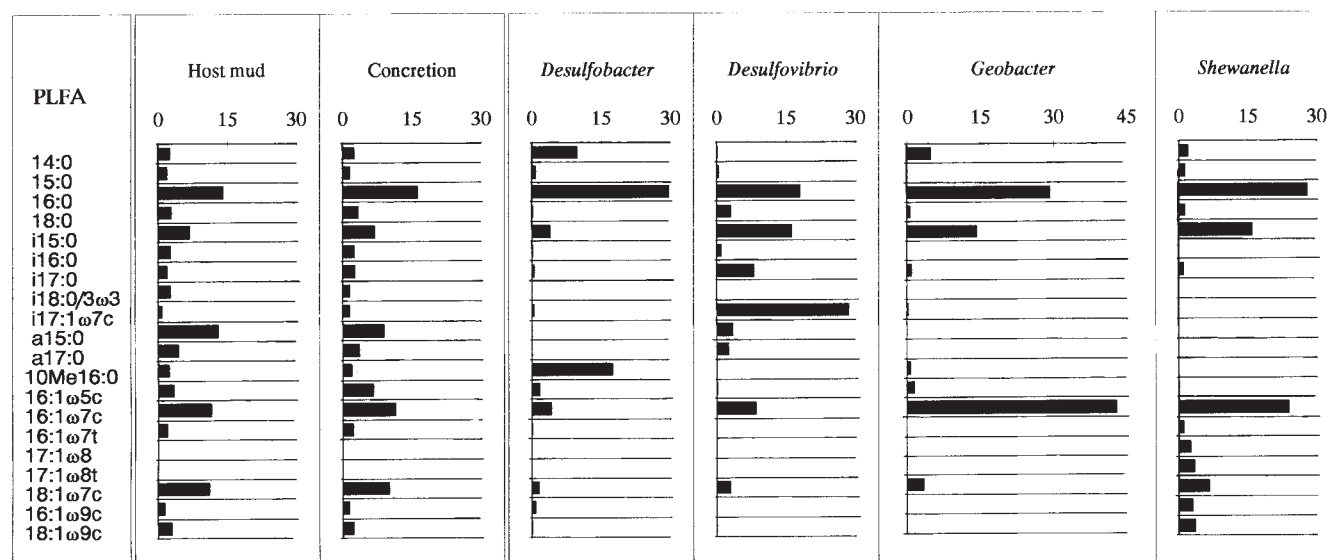


FIG. 1 Polar lipid fatty acid analyses (% of total). The significant ($>2\%$) components measured in our samples are shown, together with analyses of two SRB cultures, typical of their genera (precision, $\pm 0.3\%$). In addition to *Desulfobacter* and *Desulfovibrio*, we have included data for the dissimilatory Fe(III) reducers *Geobacter* and *Shewanella*²². The characteristic fatty acids for the two SRB examples are 10-methyl-palmitic PLFA (10 Me 16:0) and iso-17-carbon-monoenoic PLFA (i17:1 ω 7c), respectively. Some PLFA

are not generically characteristic; with starvation, some bacteria increase the proportions of the specific monoenoic PLFA with the *trans* configuration²³. Some of our samples show an increased ratio of the 16- and 18-carbon *trans* vaccenic acid to the *cis* isomer. Therefore certain of the data shown (16:1 ω 7c, 16:1 ω 7t, 18:1 ω 7c, 18:1 ω 9c) cannot be used to identify particular components of the microbial population.

similar to that found in an active subtropical estuary mud flat¹² (Table 1). The concretions had slightly lower viable biomass¹³.

Polar lipid diether (PLDE) and polar lipid tetraether (PLTE) lipids are characteristic of the archaeobacterial methanogens¹⁴. In the host mud, there was only a trace of PLTE at the minimum detectable limit and there was roughly a thousand times less PLDE than PLFA (Table 1). When compared to PLTE and PLDE analyses of several strains of methanogenic bacteria, these results suggest that the mud contained $<10^6$ methanogenic bacteria per gram. No PLTE or PLDE were detected in the concretion. These results further indicate that the formation of the siderite concretions is not associated with the activity of methanogenic bacteria.

Menaquinones (MK) are found in Gram negative anaerobes using Fe(III)¹⁵ or sulphate¹⁶ as a terminal electron acceptor as well as in Gram positive organisms, whereas ubiquinones are only found in microorganisms that can use oxygen as an electron acceptor. Fermentative eubacteria contain no respiratory quinones¹⁶. The MK/PLFA level in the concretion was two to three times higher than in the mud (Table 1). No ubiquinones were found in either the mud or the concretion. These results suggest that a high proportion of the microorganisms living in the concretion survive by anaerobic respiration such as Fe(III) or sulphate reduction.

Detailed analysis of the individual fatty acids gives more information on the types of microbial populations in the samples (Fig. 1). Some define specific components of the population, whereas others indicate nutritional status⁹. Cluster analysis of the PLFA data (Pearson's r correlation coefficient or euclidean metric) indicated the presence of two genera of sulphate-reducing bacterial (SRB), *Desulfobacter* and *Desulfovibrio*. The 10-methyl-palmitic PLFA (10 Me 16:0) is characteristic of *Desulfobacter*¹⁷ and the iso-17-carbon-monoenoic PLFA (i17:1 ω 7c) is characteristic of the *Desulfovibrio*¹⁸. The concentrations of these biomarker PLFA (Table 1) imply more *Desulfovibrio* and less *Desulfobacter* within the concretion than in the mud.

Geobacter metallireducens (formerly, strain GS-15) and *Shewanella putrefaciens* are the only well characterized organisms known to couple the oxidation of fermentation acids and/or hydrogen to the reduction of Fe(III)¹. The characteristic PFLA data gave no indication that either of these organisms was present in the mud or the concretion.

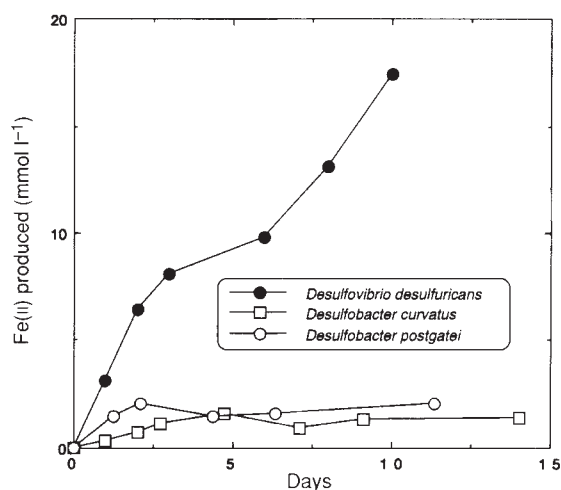
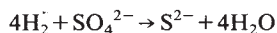
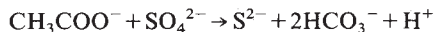


FIG. 2 Laboratory reduction of poorly crystalline Fe(III) oxide by *Desulfovibrio desulfuricans*, but not by *Desulfovibrio postgatei* nor *Desulfobacter curvatus*. The inocula for all organisms were grown in a medium with sulphate as the electron acceptor and either H₂ (*D. desulfuricans*) or acetate (*D. postgatei* and *D. curvatus*) as the electron donor. Sulphate-grown cultures were used to inoculate (10% inoculum) similar medium in which the sulphate was replaced by a synthetic, poorly crystalline Fe(III) oxide (100 mmol l⁻¹)²⁴ and the production of Fe(II) was measured over time²⁵. The means of duplicate determinations on a representative of triplicate cultures are shown (precision, ± 1 mmol l⁻¹). In the Fe(III)-containing media, the initial sulphate concentrations were 320, 740 and 440 μ M for *D. desulfuricans*, *D. postgatei* and *D. curvatus*, respectively.

This, together with the *Desulfovibrio* enrichment, led us to investigate the possibility that SRB might be involved in Fe(III) reduction. In aquatic sediments, H₂ is likely to be the most important electron donor for *Desulfovibrio*



whereas acetate is the most environmentally significant electron donor for *Desulfovibrio*



A pure culture of *Desulfovibrio desulfuricans* readily reduced Fe(III), whereas *Desulfohalobacter postgatei* and *Desulfohalobacter curvatus* did not (Fig. 2). The reduction of Fe(III) in the presence of *D. desulfuricans* could not be the result of indirect reduction of Fe(III) by sulphide because the initial sulphate concentration

of only 320 µM could provide enough sulphide to reduce less than 5% of the Fe(III) that was reduced. Furthermore, all of the sulphate was reduced within the first day of the experiment whereas most of the Fe(III) reduction was after the first day. Several other species of *Desulfovibrio* also reduce Fe(III) oxide, and the Fe(III)-reducing enzyme has now been purified from *D. vulgaris* (D.R.L., E. J. P. Phillips, J. C. Woodward and E. E. Roden, unpublished data).

To gain insight into the possible metabolism of Fe(III) and sulphate by *D. desulfuricans*, we investigated Fe(III) and sulphate reduction under limiting and nonlimiting H₂ concentrations (Fig. 3). Under nonlimiting conditions, sulphate and Fe(III) were reduced simultaneously (Fig. 3a). With sulphate as the electron acceptor, however, *D. desulfuricans* could not metabolize H₂ below ~10⁻⁵ atm (in accordance with previous results¹⁹) whereas with Fe(III), H₂ was metabolized below 3 × 10⁻⁶ atm (Fig. 3b). This suggests that at the low concentrations of H₂ found in aquatic sediments, Fe(III) may be the preferred electron acceptor.

These findings suggest that the enrichment of *Desulfovibrio* within the siderite concretion may be an important feature, controlling growth of the concretion. Oxidation of H₂ coupled to reduction of Fe(III) (reaction 1b) favours the precipitation of siderite through the formation of Fe²⁺ and OH⁻. The reaction does not produce bicarbonate. This is consistent with concretion formation in the marsh. Comparing the isotopic signature of concretion carbonate (-5.9‰) with that from organic matter oxidation (-25‰) and carbonate in sea water (0‰) suggests that most (~75%) of the carbonate comes from sea water. The additional bicarbonate (15%), derived from organic matter oxidation, could be due to the metabolism of as yet unknown Fe(III)-reducing microorganisms, or, given that the concretions contain some FeS, to the activity of SRB like *Desulfovibrio*. We are currently investigating the reasons for the apparent localized, enhanced Fe(III) reduction by *Desulfovibrio* in the zone of the concretions.

SRB may be important catalysts for Fe(III) reduction in other sedimentary environments. They are abundant in the Fe(III) reduction zones of deep aquifers in the Atlantic Coastal Plain of the USA^{20,21}. As there is no apparent sulphate reduction in these zones and the SRB cannot have been recently introduced to these deeply isolated environments, we speculate that the SRB must survive by reducing Fe(III). More intensive study of the microbiology of siderite concretions and other Fe(III)-reducing environments is needed.

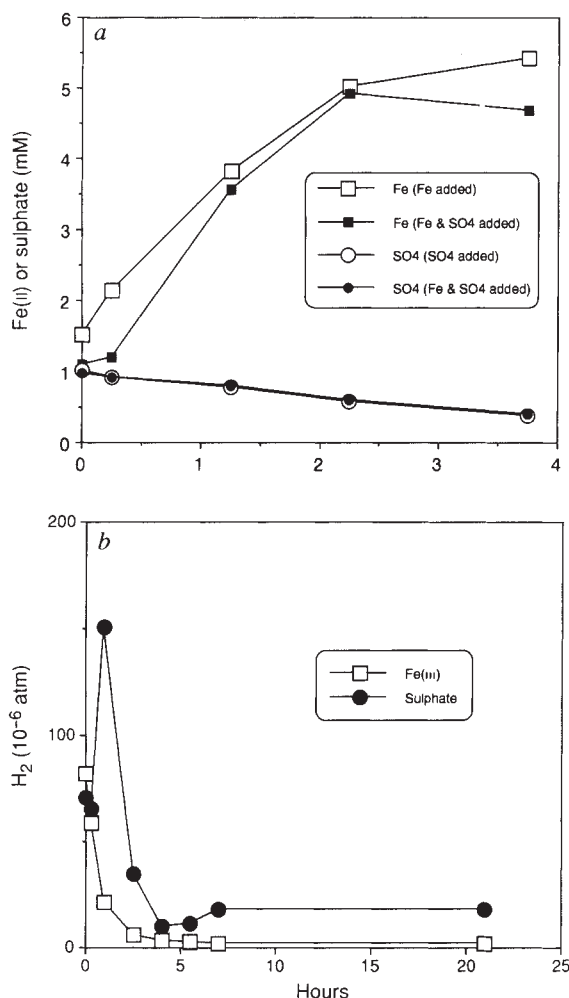


FIG. 3 H₂ oxidation coupled to Fe(III) reduction in *Desulfovibrio desulfuricans*. Under conditions of nonlimiting H₂ concentrations (a), *D. desulfuricans* reduced Fe(III) and sulphate simultaneously at rates comparable to Fe(III) and sulphate reduction when only one of the electron acceptors was provided. However, H₂ was metabolized at lower concentrations with Fe(III) than with sulphate (b).

METHODS. *D. desulfuricans* grown on H₂-sulphate medium were anaerobically collected by centrifugation and washed and re-suspended in 10 ml of anaerobic bicarbonate buffer (NaHCO₃, 25 mg; cysteine-HCl, 2.5 mg; pH 6.7) under N₂-CO₂ (80:20) in sealed 25-ml serum bottles. Cell protein concentrations were (a) 0.4 and (b) 0.2 mg ml⁻¹. H₂ was added to provide (a) 0.67 or (b) 10⁻⁴ atm. Where noted, sulphate and/or Fe(III) complexed with nitrilotriacetic acid were added from anaerobic stocks to provide initial concentrations of ~1 and 5 mM. Fe(II) and sulphate concentrations in aliquots were measured with ferrozine²⁵ and ion chromatography (precision ±1 mmol l⁻¹). H₂ was measured with a reduction gas analyser (precision, ±5 × 10⁻⁷ atm.).

Received 11 August 1992; accepted 4 January 1993.

- Lovley, D. R. *Microbiol. Rev.* **55**, 259-287 (1991).
- Berner, R. A. *Geochim. cosmochim. Acta* **48**, 605-615 (1984).
- Canfield, D. E. *Geochim. cosmochim. Acta* **53**, 619-632 (1989).
- Sørensen, J. *Appl. envir. Microbiol.* **43**, 319-324 (1982).
- Pye, K., Dickson, J. A. D., Schiavon, N., Coleman, M. L. & Cox, M. *Sedimentology* **37**, 325-343 (1990).
- Berner, R. A. *Geochim. cosmochim. Acta* **32**, 477-483 (1968).
- Gautier, D. L. *J. sedim. Petrol.* **52**, 859-871 (1982).
- Irwin, H., Curtis, C. & Coleman, M. *Nature* **269**, 209-213 (1977).
- White, D. C. *Adv. Limnol.* **31**, 1-18 (1988).
- White, D. C., Bobbie, R. J., King, J. D., Nickels, J. & Amoe, P. in *Methodology for Biomass Determinations and Microbial Activities in Sediments*, ASTM STP 673 (eds Litchfield, C. D. & Seyfried, P. L.) 87-103 (American Society for Testing and Materials, Philadelphia, 1979).
- Balkwill, D. L., Leach, F. R., Wilson, J. T., McNabb, J. F. & White, D. C. *Microb. Ecol.* **16**, 73-84 (1988).
- Findlay, R. H., Trexler, M. B., Guckert, J. B. & White, D. C. *Mar. Ecol. Prog. Ser.* **61**, 121-133 (1990).
- Vestal, J. R. & White, D. C. *Bioscience* **39**, 535-541 (1989).
- Hedrick, D. B., Guckert, J. B. & White, D. C. *J. Lipid Res.* **32**, 656-666 (1991).
- Lovley, D. R. *et al. Arch. Microbiol.* (in press).
- Hedrick, D. B. & White, D. C. *J. microbiol. Methods* **5**, 243-254 (1986).
- Dowling, N. J. E., Widdel, F. & White, D. C. *J. gen. Microbiol.* **132**, 1815-1825 (1986).
- Edlund, A., Nichols, P. D., Roffey, R. & White, D. C. *J. Lipid Res.* **26**, 982-988 (1985).
- Cord-Ruwisch, R., Seitz, H.-J. & Conrad, R. *Arch. Microbiol.* **149**, 350-357 (1988).
- Chapelle, F. H. & Lovley, D. R. *Groundwater* **30**, 29-36 (1992).
- Murphy, E. M. *et al. Water Resources Res.* **28**, 723-740 (1992).
- Lovley, D. R., Chapelle, F. H. & Phillips, E. J. P. *Geology* **18**, 954-957 (1990).
- Guckert, J. B., Hood, M. A. & White, D. C. *Appl. envir. Microbiol.* **52**, 794-801 (1986).
- Lovley, D. R. & Phillips, E. J. P. *Appl. envir. Microbiol.* **51**, 683-689 (1986).
- Lovley, D. R. & Phillips, E. J. P. *Appl. envir. Microbiol.* **54**, 1472-1480 (1988).

ACKNOWLEDGEMENTS. We thank D. B. Ringelberg for cluster analysis, and J. Woodward and E. Phillips for technical assistance.

Mechanisms of magma eruption and emplacement at Mt Etna between 1989 and 1992

Hazel Rymer*, John B. Murray*, Geoff C. Brown*§, Fabrizio Ferrucci† & William J. McGuire‡

* Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

† Gruppo Nazionale Vulcanologia, Centro Operativo Catania, 6 Viale Regina Margherita, 95123 Catania, Italy

‡ Cheltenham & Gloucester College of Higher Education, Shaftesbury Hall, St Georges Place, Cheltenham GL50 3PP, UK

THE present eruption sequence of Mt Etna began during September 1989 with extensive fracturing, and culminated in 1991–93 with the largest lava eruption this century¹. Here we present measurements of gravity and ground deformation which, in conjunction with seismic data, provide a detailed record of the processes occurring before and during eruption. These reflect a complex interplay between magma injection into the paths of lowest compressive stress through the volcano flanks, forcing elastic deformation, and fracture growth in response to gravitational stresses. We find that after the end of the eruptive activity in 1989, fractures in the south-southeast flank of the volcano remained unfilled until some months before the 1991–93 eruption. These results indicate the utility of microgravity monitoring, as the precursory processes that we identify were largely undetected by seismic and ground deformation monitoring.

At a late stage during the explosive activity from the southeast crater (10 September–6 October 1989), two prominent fissures developed, one erupting to the northeast² and the second, a non-erupting fissure, extending 7 km towards the south-southeast. Seismic^{3,4}, soil gas emanations⁵, thermal imaging⁶, gravity⁷ and our own deformation and gravity data indicate that magma was not emplaced within the latter fissure south of station TDF (Fig. 1). After a 2-year period of quiescence, explosive activity near the summit along the same SSE fissure on 14 December 1991 heralded the start of the present eruption, which has continued to produce basaltic magma to the present day from further down the fissure at an elevation of 2,400 m above sea level (a.s.l.). The main 1991–93 eruptive vent is located where the intrusion intersects the topography, 3 km from the summit, in a region where the theoretical free surface of magma is higher than the topography⁸.

The 1989–90 period was marked by initial gravity increase (up to 100 μ Gal) at stations near the summit (June–November 1989) followed by decreases of similar magnitude (November 1989–June 1990). At the same time, elevation data indicate progressive deflation of the summit region (Table 1). There is no obvious correlation between these data sets, and the changes in height are much smaller than required to explain the variations in gravity. Subsurface mass increases of 10^9 – 10^{10} kg followed by decreases must have occurred, and these bracketed the January 1990 summit pyroclastic eruption which ejected 10^9 kg of material. We suggest that magma was emplaced above station TDF before January 1990 and that this summit eruption reduced magma levels.

Between June 1990 and June 1991, there was no eruptive activity although, as usual, the summit craters were degassing freely. Gravity data indicate 400- μ Gal increases at summit stations and a zone trending south-southeast in which gravity was increased by 100 μ Gal (Table 1 and Fig. 2a). Again, height changes over this period (Table 1 and Fig. 2b) are small, so a linear zone of subsurface mass increases connected to a central

feeder conduit is inferred. The south-southeast-trending increase is most probably the result of a shallow magma intrusion. From the relative positions of the 1989 fracture and the gravity high (Figs 1 and 2), it appears that the intrusion is either vertical and about 100 m west of the fracture, or inclined downwards west away from the fracture. Magma intrusion was essentially aseismic; several earthquake sequences occurred between February and May 1991, but the sources were deep (7–20 km). Surface deformation was small and not of the form expected (inflation flanking the intrusion) when magma is emplaced elastically⁹. We therefore propose that magma migrated passively into a pre-existing subsurface fracture some time between June 1990 and June 1991.

By October 1991 gravity had fallen again (Table 1) at summit stations, but values were still higher than in June 1990. A smaller gravity decrease broadly following the south-southeast fissure was accompanied by minor inflation (Table 1). EDM network (Fig. 1) line length increases also indicate tumescence in the

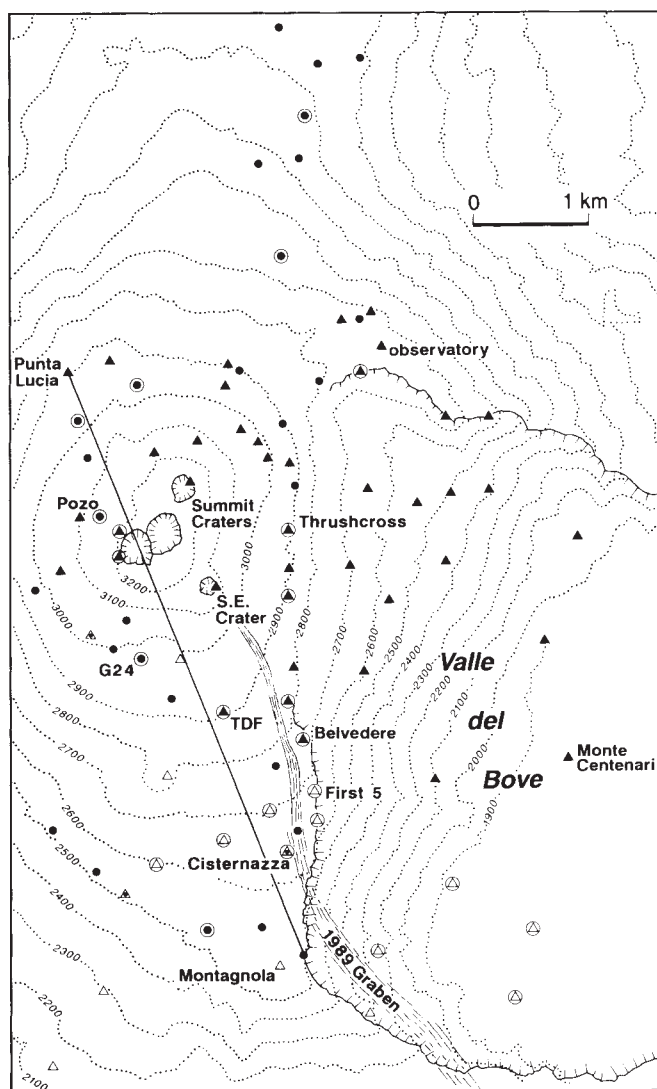


FIG. 1 Location map for summit ground deformation and gravity surveys. The trend of the fracture or graben left after the 1989 eruption is shown. The eruptive vent for the 1991–93 activity lies just below the intersection of the 1989 fracture and the cliff wall of the Valle del Bove. Principal levelling stations are indicated by black circles, triangulation stations are black triangles, slope distances only are measured at stations marked by open triangles and gravity stations are indicated by large open circles. The profile marked is used in cross-section in Fig. 3.

§ Geoff Brown died tragically in the eruption of Galeras volcano²⁰ on 14 January 1993.

region above the future eruption site. Deep (4–5 km) low-energy microseismicity was observed continuously in this area between July and August 1991. The combination of seismic, gravity and ground deformation evidence suggests vesiculation and expansion of frothy magma and/or downwards fracture propagation and magma seepage. Pre-eruption vesiculation has been inferred previously using microgravity at andesitic stratocones (Poas and Pacaya)^{10–12} where only 1–2% change in the degree of vesiculation accounts for the observed changes (of $\sim 100 \mu\text{Gal}$). At Etna, highly vesiculated basaltic magma is usually erupted, so magma vesiculation before eruption is a plausible explanation of the gravity decrease. On 14 December 1991, the south-southeast-trending fracture widened (maximum extension by April 1992, 1.5 m), lava was erupted close to the summit, and by 15 December activity had moved down to a site 2,400 m a.s.l. where the fracture intersects the western cliff wall of the Valle del Bove¹. The onset of the eruption was accompanied by intense seismic activity 1–3 km beneath the fracture between the eruptive boccas and the summit. After a stronger sequence, observed on 15 December and related to the elastic response of the edifice to the intrusive body¹³, seismicity returned to background levels, then virtually vanished.

Gravity increases occurred during the eruption and, by June 1992, gravity at stations on the south flank was similar to but higher than in June 1991 (Table 1). However, substantial eleva-

tion decreases occurred (Table 1) at stations along the fracture. After allowing for this subsidence there are no significant residual gravity changes between June 1991 and June 1992, except at the summit crater where there was a decrease of $80 \mu\text{Gal}$. At the same time, up to 37 cm of inflation occurred 1 km east of the fracture and 7 cm of inflation 0.5 km west of it, suggesting that elastic deformation occurred at the onset of the eruption, in response to increased magma pressure. Dynamic equilibrium within the fracture system was apparently maintained until 14 December 1991, when the magma pressure exceeded the lithostatic stresses and the eruption began briefly at the summit end of the fracture, then continued at a bocca further down. At this time, the more common elastic deformation and further dyke intrusion occurred.

The profile marked on Fig. 1 follows the largest June 1990–June 1991 gravity changes and the south-southeast fracture trend. Modelling indicates an increase in magma level within the summit feeder conduit, and within an open fracture trending south-southeast (Fig. 3). The latter was not a forceful intrusion, because local elevation changes were far too small and the lack of seismicity implies that magma did not force the surrounding rock apart. We propose instead that magma exploited the 1989 fracture system which remained partly void until new magma filled it between June 1990 and June 1991. Buried fractures of this type are believed to occur as a natural consequence of

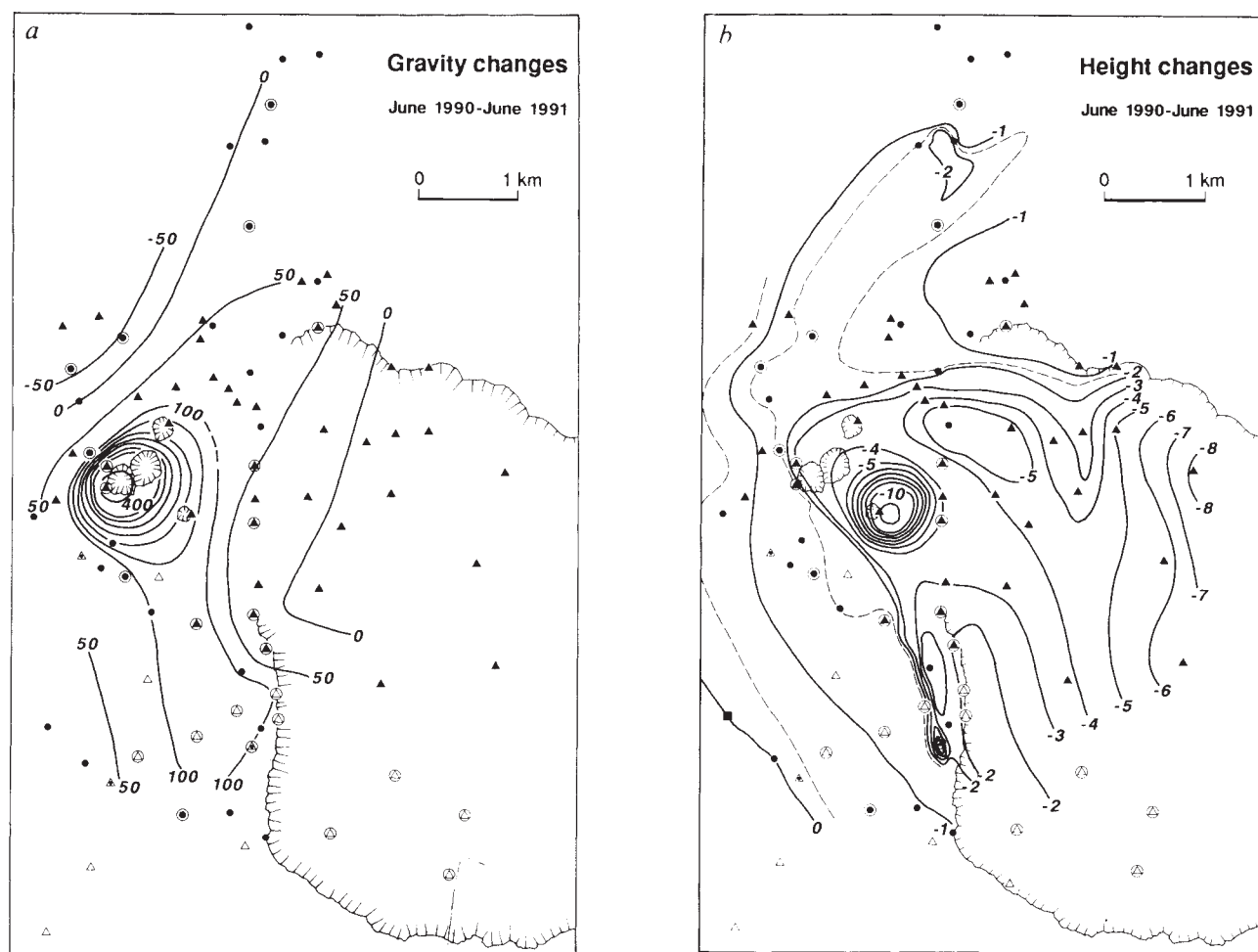


FIG. 2 Contour maps showing (a) gravity changes (μGal) and (b) height changes (cm) observed between June 1990 and June 1991 (data from Table 1). Gravity data are expressed relative to a base 5 km from the summit which has proved to be gravitationally invariant with respect to other stations further down the volcano flanks and off the volcano. The deformation data reference is at station Piccolo Rifugio (the gravity station between Cister-

nazza and Montagnola). For June 1990 to June 1991, gravity increases occur at all stations but are largest at the summit. A south-southeast-trending gravity high follows the trend of the 1989 fracture system and is displaced ~ 100 m west. Height changes indicate subsidence of the southeast crater and a small deflation along the strike of the south-southeast trending fracture.

TABLE 1 Gravity and height data

Gravity change (mGal)							
Station	June 1989	Nov 1989	June 1990	June 1991	Oct 1991	April 1992	June 1992
TDF	-0.074	0.000	-0.103	0.024	-0.029	0.145	0.039
Piccolo			-0.058	0.005	-0.063		-0.004
Cisternazza	0.000	0.000		0.020	-0.040	0.006	0.061
Belvedere	0.042	0.000	0.032	0.054	-0.002	0.143	0.205
G24	-0.106	0.000	-0.075	0.016	-0.048		-0.008
Pozo		0.000	0.000	0.408	0.109	0.378	0.331
Thrushcross		0.000	0.021		0.033		0.040
First 5		0.000	0.186	0.286	0.211	0.403	0.647

Height change (m)							
Station	June 1989	Nov 1989	June 1990	June 1991	Oct 1991	April 1992	June 1992
TDF	0.037	0.000	-0.001	-0.005	0.000	-0.060	-0.056
Piccolo	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cisternazza	0.008	0.000	-0.013	-0.064	-0.062	-0.020	-0.102
Belvedere	-0.165	0.000	-0.017	-0.025	-0.022	-0.520	-0.584
G24	0.033	0.000	0.005	0.001	0.007		-0.015
Pozo		0.000	0.000	0.021	-0.021	-0.010	0.013
Thrushcross	0.031	0.000	0.000	-0.024	-0.021		0.096
First 5		0.000	-0.028	-0.030	-0.028	-1.180	-1.175

Summary of key gravity (Δg) and height (Δh) data for the period June 1989 to June 1992. Gravity data (in mGal) are expressed relative to the reference station at the Rifugio Sapienza 5 km from the summit of Etna. The reference point for the height data (m) is Rifugio Piccolo (between Cisternazza and Montagnola) and values for November 1989 are set to zero, after the SSE-trending fractures had formed and Strombolian activity had ceased. The standard deviation on the gravity data is less than 30 μ Gal and on the elevation data it is less than 1 cm.

within-edifice strain at intraplate volcanoes¹⁴, and void regions have been postulated within the Kilauean rift zone^{15,16}. Here, with several closely spaced data we are able to model a previously void region (Fig. 3) and find that it extends at least 3 km laterally and 1 km vertically. The model is most sensitive to the lateral extent of the dyke and to its upper surface. Because no seismic sources have been detected above sea level in the whole period of observation, we conclude that the fractured region and possibly the intrusion also extend down below sea level.

The suggestion that Etna fissures may store magma before eruption was originally made¹⁷ to explain small gravity changes accompanying drainage of a dyke in March 1981. In the 1989

case, magma was forced into two arms of an underground fissure network (northeast and south-southeast of the southeast crater) but may have been stopped by a structural boundary (the caldera del Piano) seen on the surface as a break in slope and encompassing the southeast crater and stations Belvedere and TDF (Fig. 1). The most likely reason for continued propagation of the 1989 fissure, although in a slightly different direction beyond the postulated boundary, is gravitational slope failure on the western wall of the Valle del Bove¹⁸. The minimum volume of the fissure, derived from the model of the intrusion, is 10^7 m³.

Although the 1991 intrusion was emplaced aseismically, low-energy microseismicity between June and December 1991 suggests that vigorous degassing continued after emplacement. Strombolian activity in the summit craters was vigorous, with spatter and bombs thrown 200 m into the air. We infer that the intrusion was actively degassing, the frothy material remaining mobile and hot. Because much of the erupted magma was gas-poor, we propose that a subsurface connection exists between the intrusion and the summit feeder conduit (Fig. 3) allowing gas to escape at the summit craters. A single deep connection to the magma supply maintains the heat input by convection.

The increase in gravity once the eruption had begun indicates less degassing and that the fractured region contained a high level of slow-moving volatile-poor magma. Summit gravity increases were smaller, suggesting either that frothy magma resided in the main feeder conduit or that the magma level was much lower than before the eruption.

We have shown that fissure networks within a volcanic edifice can remain open between eruptions. Although such fractures may not reach the surface, their effect is to reduce the pressure of incoming magma from magmatic towards atmospheric. Thus vigorous magmatic degassing within the edifice is facilitated and the probability of explosive eruption is decreased. Fractures on Etna occur in well defined zones related to the asymmetric topography (and hence gravitational instability¹⁹) and the regional tectonic regime. Once formed, these fractures may remain open for months or years providing low stress paths where magma may intrude passively with little seismic or ground deformation; only gravity changes can detect the mass increase. If the intrusion does not have time to freeze before subsequent or continued injection, then the 'lubricated' region may channel magma towards the eventual eruption site. A test for this model will be to observe the relationship between horizontal extension (fracture formation), gravity changes (mass changes) and subsequent eruption at dyke-forming volcanoes. □

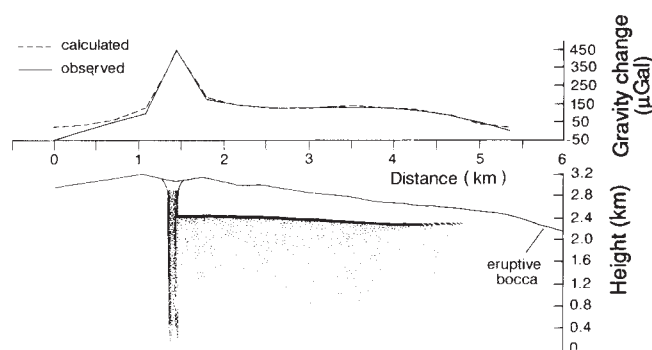


FIG. 3 Computed gravitational effect from a 2.5D model of the 1990–91 changes in magma level within the summit feeder and the proposed fissured region beneath the upper southern flank (located on Fig. 1). The density of the intrusions is assumed to be $2,800 \text{ kg m}^{-3}$, the same as the surrounding material. Note that the total strike width of the feeder pipe is 50 m and the width of the dyke 4 m. The model does not suggest that these are the limits of the bodies, simply that these are the sort of changes between June 1990 and June 1991 which would fit the observed gravity data. Indeed the feeder pipe almost certainly extends down to a large magma reservoir well below sea level. Seismic data indicate sources at 2 km depth, which roughly coincides with the modelled base here. The approximate location of the main eruptive vent for the 1991–92 activity is also shown, although in fact it is slightly to the east of this profile (into the page).

Received 12 October; accepted 29 December 1992.

- SEAN Bull. **17**, 12–16 (1992).
- Bertagnini, A. et al. in *Mt Etna: the 1989 Eruption* (eds Barberi, F. & Bertagnini, A. & Landi, P.) 10–22 (Giardini, Pisa, 1990).
- Ferrucci, F. in *Mt Etna: the 1990 Eruption* (eds Barberi, F., Bertagnini, A. & Landi, P.) 36–43 (Giardini, Pisa, 1990).
- Ferrucci, F. et al. *J. Volcan. geotherm. Res.* (in the press).
- Chiodini, G. et al. in *Mt Etna: the 1990 Eruption* (eds Barberi, F., Bertagnini, A. & Landi, P.) 65–67 (Giardini, 1990).
- Bianchi, R. et al. in *Mt Etna: the 1990 Eruption* (eds Barberi, F., Bertagnini, A. & Landi, P.) 69–72 (Giardini, 1990).
- Budetta, G., Grimaldi, M. & Luongo, G. in *Mt Etna: the 1990 Eruption* (eds Barberi, F., Bertagnini, A. & Landi, P.) 56–57 (Giardini, 1990).
- Bonafede, M. & Boschi, E. *J. Volcan. geotherm. Res.* **49**, 249–263 (1992).
- Murray, J. B. in *Magma Transport and Storage*, Ch. 17 (ed. Ryan, M. P.) (Wiley, Chichester, 1991).
- Eggers, A. A. *J. Volcan. geotherm. Res.* **19**, 223–237 (1983).
- Rymer, H. & Brown, G. C. *Nature* **311**, 243–245 (1984).
- Rymer, H. & Brown, G. C. *Bull. volcan.* **49**, 389–398 (1987).
- Ferrucci, F. & Paterne, D. *J. Volcan. geotherm. Res.* (in the press).
- Chevallier, L. & Verwoerd, W. J. *Bull. volcan.* **52**, 413–425 (1990).
- Dzurisin, D. et al. *J. Volcan. geotherm. Res.* **7**, 241–269 (1980).
- Parfitt, E. *J. geophys. Res.* **96**, 10101–10112 (1991).
- Sanderson, T. J. O. *Nature* **297**, 487–490 (1982).
- McGuire, W. J., Murray, J. B. & Pullen, A. D. *J. geol. Soc.* **148**, 577–584 (1991).
- McGuire, W. J. & Pullen, A. D. *J. Volcan. geotherm. Res.* **38**, 325–344 (1989).
- Nature* **361**, 193 (1993).

ACKNOWLEDGEMENTS. The Royal Society (London), The European Commission, Earthwatch (Belmont, Mass.) and the Gruppo Nazionale per la Vulcanologia are gratefully acknowledged for financial support. Comments from J. Fink improved the manuscript.

New Silurian and Devonian fork-tailed 'thelodonts' are jawless vertebrates with stomachs and deep bodies

Mark V. H. Wilson & Michael W. Caldwell*

Department of Zoology and Laboratory for Vertebrate Paleontology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

* Present address: Redpath Museum, McGill University, 859 Sherbrooke Street W., Montréal, Québec H3A 2K6, Canada

ALL agnathans (jawless vertebrates) are generally considered to have lacked a stomach, as do the living lampreys. Thelodonts are agnathans of Silurian and Devonian age whose bodies were covered by tiny, hollow, tooth-like scales which are useful for correlating rocks. We have discovered a new group of agnathans with 'thelodont' scales but with a body plan very different from that of previously known thelodonts. Sediment infillings reveal gut morphology that includes a probable stomach, suggesting that stomachs arose before jaws in vertebrate evolution. These new fork-tailed 'thelodonts' are also the first agnathans known to have deep, compressed bodies. They represent an important new clade of early vertebrates and are potential close relatives of gnathostomes (jawed vertebrates).

The relationships of thelodonts as previously known have been controversial: they were considered to be relatives of pteraspidiiforms, anaspidiiforms, gnathostomes or chondrichthyans^{1,2}, or to be among the most primitive vertebrates^{2,3}. Experts also disagree about their monophyly^{4,5}. Scale structure indicated two groups, one including *Lanarkia* and *Turinia*^{6,7} and the other including *Loganellia* and *Phlebolepis*^{8,9}. But articulated skeletons for both scale groups have a wide, depressed

anterior body and head, small eyes, a long slender tail, dorsal, anal and anterior paired fins, a branchial row that is under the paired fins and not obviously oblique, and a hypocercal tail. The similar body form has been interpreted as evidence that thelodonts are monophyletic and has been supported by a synapomorphy of the scale base².

New thelodont discoveries reported here are in undoubted marine rocks. One species occurs with pteraspidiiforms in an early Wenlockian (420 Myr BP; Silurian) horizon in section AV1 of the Road River Formation^{10,11} (near Avalanche Lake, NWT, Canada). Four species occur 40 km NW along with diverse agnathans and acanthodians^{12,13} in an early Devonian horizon (410 Myr BP) of the Delorme Group.

Like previously known thelodonts, the new fossils all have bodies covered in characteristic 'thelodont' scales (Fig. 1), each with a dentinous crown, a bony base, and a single large pulp cavity². Scales vary in shape on different parts of the body. Both we and others¹² originally interpreted the first specimens as conventional thelodonts with depressed bodies, but new specimens show that they are very different. All new taxa, constituting a group of approximately ordinal rank, will be described elsewhere. Until then, we suggest the term 'fork-tailed thelodonts'.

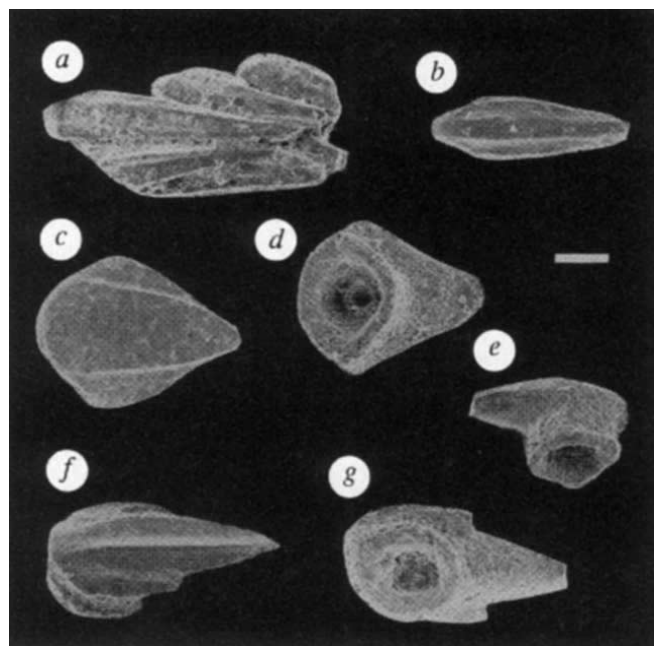


FIG. 1 Scanning electron micrographs of scales of new fork-tailed thelodonts, each scale removed from an articulated specimen. *a, b*, Crown view of flank scales of the Silurian species from specimen (University of Alberta Laboratory for Vertebrate Paleontology UALVP) 33000: *a*, group of scales; *b*, single scale. *c, d* and *e*, Flank scales of first Devonian species, UALVP 32951: *c*, crown view; *d*, root view; *e*, side view. *f, g*, Near mid-ventral scales of fourth Devonian species, UALVP 19338: *f*, oblique crown view; *g*, root view. Scale bar, 100 μ m.

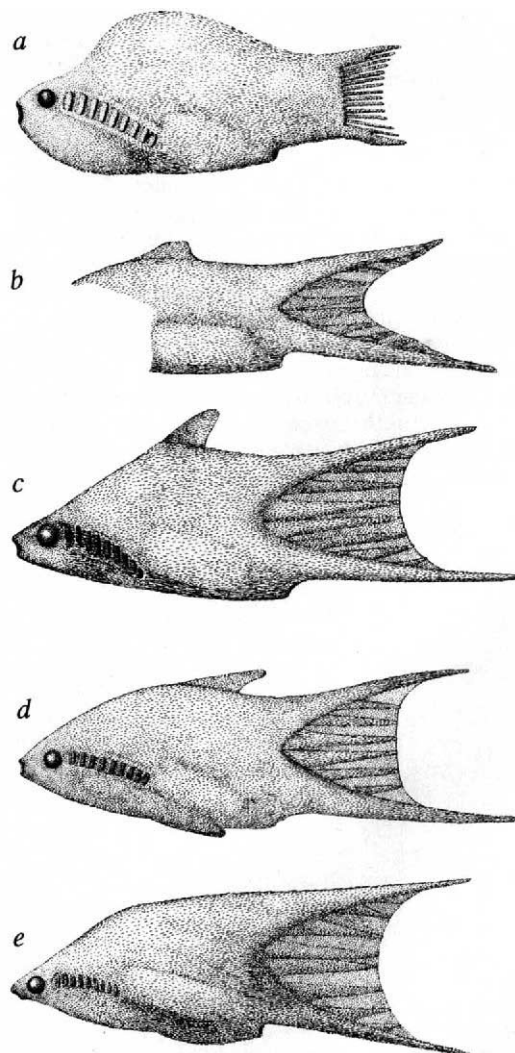


FIG. 2 Skeletal reconstructions of new fork-tailed and deep-bodied Silurian and Devonian thelodonts. Mouth shape is not definitely known. *a*, Silurian species known from 20 specimens. *b*, First Devonian species known from two specimens. *c*, Second Devonian species known from five specimens. *d*, Third Devonian species known from five specimens. *e*, Fourth Devonian species known from 17 specimens.

Species are easily distinguished by scale shape (Fig. 1), body-size and shape, presence and position of dorsal fin, number of caudal lobes, and position of anus (Figs 2 and 3). All specimens are laterally flattened. Within a species all are of similar size, suggesting determinate growth. The body has a bulbous visceral part (pharynx plus body cavity) and a deep, compressed somatic part (trunk plus tail) with a hump-backed shape. The tail is supported by 8–14 tapered, scaled lobes, connected in Devonian species by a scaled web. The large lower and upper lobes give the tail a nearly symmetrical, forked shape. An anal fin is absent but some have a dorsal fin. No paired fins are seen above the branchial row, but some species have ventral paired fins difficult to homologize with paired fins of other agnathans. The branchial row is oblique, condensed, and slightly curved. Its 8 to 9 openings are separated by rows of composite scales, each with a long base joining rows of tiny crowns (similar branchial scales occur in *Loganellia*¹⁴). Eyes are large and just in front of the first branchial opening. Gut endocasts reveal a large, oblique, barrel-shaped organ that we interpret as a stomach, as well as a short intestine that ends behind an anal prominence.

The thelodont ?*Canonia heintzeae* (Dineley and Loeffler)^{2,7,12}, originally described as flattened dorsoventrally with 'pectoral' fins like conventional thelodonts, is actually a laterally flattened, fork-tailed thelodont, though probably a distinct species. The dorsal fin, ventral fins, intestine, anal prominence, and apex of the tail fork can all be seen in the original figure.

Zoologists have often theorized that all jawless vertebrates lacked a stomach because one is absent in cephalochordates, hagfishes, and lampreys, and because filter-feeders are thought not to need one^{15–19}, making presence of a stomach a gnathostome synapomorphy thought to be associated with jaws and eating large prey. Intestines in agnathans are also assumed to be simple, short, and straight. Although the internal structure of the head is well preserved in cephalaspidiforms and galeaspidiforms^{20,21}, internal structures of the trunk are rarely seen. Some anaspids and fossil lampreys have traces of abdominal organs preserved: for example, dark areas in *Endeiolepis*, *Mayomyzon* and *Gilpichthys* have been interpreted as livers^{22,23}, dark streaks as intestines²⁴, and a possible stomach has been noted in *Endeiolepis*²⁵. Traces of a gut have been seen in some

thelodont species² and we note that the holotype specimen of the thelodont *Turinia* also has an area of mud behind the branchial region⁷. However, potential stomachs have either been overlooked or found unconvincing.

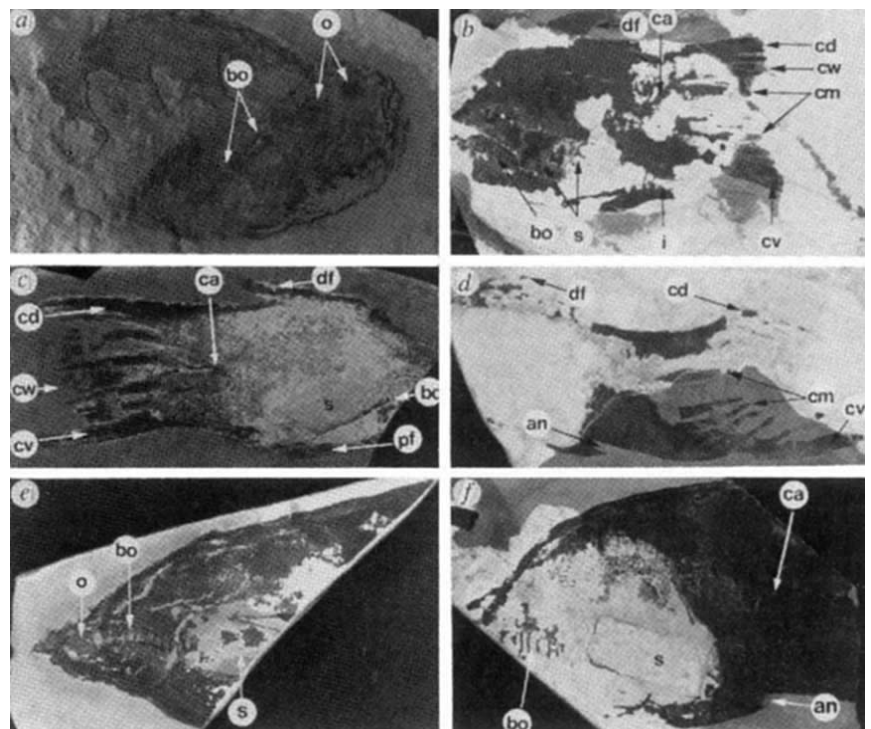
In contrast, the large, oblique, barrel-shaped organ of the fork-tailed thelodonts (Fig. 3) is unmistakable. We think it is a stomach rather than a large intestine because of its discrete shape and because there is evidence for a pyloric valve: in some specimens both stomach and intestine are filled with sediment, whereas in others only the stomach is filled (Fig. 3). We suggest that a stomach evolved in vertebrates before jaws, that it occurred in animals with short, conical intestines, and that it was very probably present in remote ancestors of lampreys also.

The compressed bodies and large tails of the fork-tailed thelodonts point to active swimming²⁶, and their large eyes and presumed stomach are consistent with predatory habits. The predaceous model implies that vertebrates might have been able to swallow large prey before the origin of jaws. However, the muddy gut fillings indicate suspension feeding, as has been suggested for the Devonian placoderm *Bothriolepis*²⁷ and some anaspidiforms²⁸. We currently favour suspension feeding for the new group, concluding that evolution of a stomach could well have preceded the need to store and digest large prey.

Scales of our Devonian species resemble those of Devonian thelodonts in the families Apalolepididae and Nikoliviidae²⁹, yet our Silurian species has scales more like those of *Loganellia* (Fig. 1). But body form proves that the fork-tailed thelodonts are closely related to each other and distinct from conventional thelodonts and other agnathans. The unusual tail, the compressed, oblique, postorbital branchial row, and the thelodont scales might be considered as good indicators of relationships, yet one (tail shape) is shared with pteraspids^{11,30}, the second (branchial row) with anaspidiforms⁴, and the third (scale morphology) with conventional thelodonts², groups that are thought to be distantly related.

Dorsal, anal and paired fins, heterocercal tails, and micromeric skeletons have all been said to indicate close relationship of various agnathans with gnathostomes^{1,2,4}. The fork-tailed thelodonts also share many features with gnathostomes (large eyes, micromeric skeletons, stomachs, and dorsal and paired

FIG. 3 Specimens and important features of fork-tailed 'thelodonts'. All specimens are preserved in lateral view. *a*, Head and body of Silurian species, UALVP 32992. *b*, Body and tail of second Devonian species, UALVP 33024, stomach and intestine filled with sediment. *c*, Body and tail of third Devonian species, UALVP 32958, stomach filled with sediment. *d*, Body and tail of first Devonian species, UALVP 33027. *e*, Head and anterior body of fourth Devonian species, UALVP 33023, stomach filled with sediment. *f*, Body and anterior tail of fourth Devonian species, UALVP 38075, stomach filled with sediment but intestine not filled. Abbreviations: an, anal notch; bo, branchial openings; ca, apex of caudal fork; cd, dorsal caudal lobe; cm, middle caudal lobe; cv, ventral caudal lobe; cw, scale-covered caudal fin web; df, dorsal fin; i, endocast of intestine; o, orbit; pf, ventral paired fin; s, endocast of stomach.



fins), and must therefore be part of any search for the closest relatives of vertebrates with jaws. □

Received 28 September; accepted 3 December 1992.

1. Forey, P. L. *J. Vertebr. Paleontol.* **4**, 330–343 (1984).
2. Turner, S. in *Early Vertebrates and Related Problems of Evolutionary Biology* (eds Mee-mann, C., Yu-hai, L. & Guo-rui, Z.) 87–120 (Science Press, Beijing, China, 1991).
3. Bleick, A. & Janvier, P. in *Murchison Symposium: Proceedings of an International Conference on the Silurian System, Special Papers in Palaeontology* **44** (eds Bassett, M. G., Lane, P. D. & Edwards, D.) 345–389 (1991).
4. Janvier, P. *J. Vertebr. Paleontol.* **1**, 121–159 (1981).
5. Novitskaya, L. I. & Karatajute-Talimaa, V. N. *Paleontol. J.* **23**, 1–14 (1989).
6. Ritchie, A. thesis, Univ. Edinburgh (1963).
7. Turner, S. *Palaeontology* **25**, 879–889 (1982).
8. Märss, T. *J. Vertebr. Paleontol.* **6**, 1–11 (1986).
9. Ritchie, A. in *Current Problems of Lower Vertebrate Phylogeny* (eds Örvig, T.) 81–88 (Almqvist & Wiksell, Stockholm, 1968).
10. Over, D. J. & Chatterton, B. D. E. *Geol. Palaeontol.* **21**, 1–49 (1987).
11. Soehn, K. & Wilson, M. V. H. *J. Vertebr. Paleontol.* **10**, 405–419 (1990).
12. Dineley, D. L. & Loeffler, E. J. *The Palaeontological Association, Special Papers in Palaeontology* **18**, 1–214 (1976).
13. Bernacke, G. M. & Dineley, D. L. *Palaeontogr., Abt. A* **158**, 25 (1977).
14. Vergoossen, J. M. J. *Geologie en Mijnbouw* **71**, 51–64 (1992).
15. Romer, A. S. & Parsons, T. S. *The Vertebrate Body* (Saunders, Toronto, 1977).
16. Kluge, A. *Chordate Structure and Function* (Macmillan, New York, 1977).
17. Walker, W. F. Jr *Functional Anatomy of the Vertebrates: An Evolutionary Perspective* (Saunders College, Philadelphia, 1987).
18. Young, J. Z. *The Life of Vertebrates* (Clarendon, Oxford, 1981).
19. Youson, J. H. in *Evolutionary Biology of Primitive Fishes* (eds Foreman, R. E., Gorbman, A., Dodd, J. M. & Olsson, R.) 59–68 (Plenum, New York, 1985).
20. Halstead, L. B. *Nature* **282**, 833–836 (1979).
21. Janvier, P. *Les Céphalaspides du Spitzberg* (Éditions du Centre National de la Recherche Scientifique, Paris, 1985).
22. Bardack, D. & Richardson, E. S. Jr *Fieldiana Geol.* **33**, 489–510 (1977).
23. Arsenault, M. & Janvier, P. in *Early Vertebrates and Related Problems of Evolutionary Biology* (eds Mee-mann, C., Yu-hai, L. & Guo-rui, Z.) 19–40 (Science Press, Beijing, China, 1991).
24. Lund, R. & Janvier, P. *Géobios* **19**, 647–652 (1986).
25. Stensio, E. A. *Kungliga Svenska Vetenskapsakademien Handlingar* (3) **18**, 1–25 (1939).
26. Mallatt, J. in *Evolutionary Biology of Primitive Fishes* (eds Foreman, R. E., Gorbman, A., Dodd, J. M. & Olsson, R.) 59–68 (Plenum, New York, 1985).
27. Denison, R. H. *J. Paleontol.* **15**, 353–361 (1941).
28. Gilmore, B. *Palaeontology* **35**, 319–333 (1992).
29. Vieth, J. *Goettinger Arb. Geol. Paläontol.* **23**, 69 (1980).
30. Frickhinger, K. A. *Fossilien Atlas: Fische* (Mergus, Verlag für Natur- und Heimtierkunde, Melle, Germany, 1991).

ACKNOWLEDGEMENTS. We thank B. Chatterton for sharing collections and field work, A. Lindoe and S. Derr for assistance, C. Gans, V. Karatajute-Talimaa, T. Märss, A. Ritchie, K. Soehn, R. Thorsteinsson and S. Turner for discussion, J. Nelson and S. Turner for reading drafts, and P. Forey and P. Janvier for helpful reviews. The work was supported by Canada NSERC and NSTP, by the University of Alberta CRF, and is a contribution to IGCP 328–Palaeozoic Microvertebrates.

Origin of underwater hearing in whales

J. G. M. Thewissen* & S. T. Hussain†

* Department of Biological Anthropology and Anatomy, Duke University Medical School, Durham, North Carolina 27710, USA
† College of Medicine, Department of Anatomy, Howard University, Washington DC 20059, USA

ALL described fossil and Recent cetaceans have relatively similar ear bones (malleus, incus and stapes) that strongly diverge from those of land mammals^{1–4}. Here we report that the hearing organ of the oldest whale, *Pakicetus*, is the only known intermediate between that of land mammals and aquatic cetaceans (whales, dolphins and porpoises). The incus of *Pakicetus* is intermediate with respect to inflation, crural proportions, and position of the malleolar joint. The incus and mandible of *Pakicetus* indicate that the path of soundwaves to its ear resembled that of land mammals. These fossils suggest that the first whale was amphibious, and corroborate the hypothesis that artiodactyls (for example, pigs, camels and ruminants) are the closest extant relatives of cetaceans.

In 1991 and 1992, we collected jaws and ear bones of the oldest known whale, the archaeocete *Pakicetus*^{5,6}, in an early Eocene river-deposited conglomerate in the Kala Chitta Hills of Pakistan (Locality 62). Here we discuss implications of these fossils for (1) the function of the ancestral whale ear and its

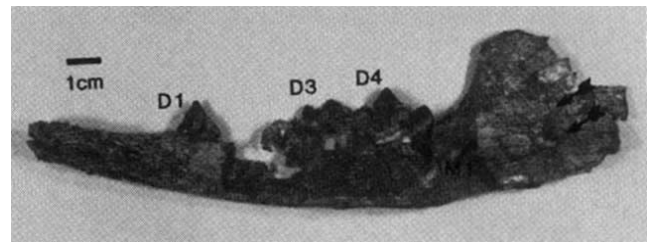


FIG. 1 Right mandible of juvenile *Pakicetus* (Howard-Geological Survey of Pakistan 91034). This is the only complete jaw known for this taxon and shows that the mandibular foramen (between arrows) is small, like that of a terrestrial mammal, but unlike that of all other cetaceans. Alveoli are preserved for I1 through D4, and crowns for D1, D3 and D4 are visible. M1 is preserved in its alveolus. As in all archaeocetes, the teeth lack crushing basins. A single cusp dominates trigonid and talonid.

evolution into the specialized hearing system of modern whales, (2) the phylogenetic ties between whales and other mammals and (3) the habitat in which the earliest whales lived.

The lower jaw of all described cetaceans has a greatly enlarged mandibular foramen⁸, which lodges (in Recent odontocetes) a fat pad that extends to the ear region (lateral tympanic)⁹. The fat pad and foramen are often assumed to be involved in underwater hearing: sound vibrations received by the lower jaw are transmitted by the fat pad to the middle ear⁹. *Pakicetus* is the only cetacean in which the mandibular foramen is small (Fig. 1), as is the case in all terrestrial mammals. It thus lacked the fat pad, and sounds reached its eardrum following the external auditory meatus as in terrestrial mammals.

The incus of all described cetaceans has a greatly inflated body and crus longum (Fig. 2), whereas *Pakicetus* shows only slight incudal inflation near the joint for the malleus (Fig. 2). Inflated earbones also occur in true seals (Phocidae^{10,11}), walrus (Odobenidae¹²) and sea cows (Sirenia¹⁰). Some physiologists^{13,14} have suggested that this may be related to a specialized system of underwater hearing based on the inertia of malleus and incus with respect to the rest of the ear. Owing to their increased mass (inertia), the malleus and incus respond more slowly to incoming (sound) vibrations than does the periotic, and this induces the stapes (lodged between the incus and the periotic) to cause standing waves in the perilymph of the periotic. These waves are perceived in hearing^{13,14}. There is, however, no consensus on the hearing mechanism of extant cetaceans, and lever-based mechanisms similar to those of terrestrial mammals might also be present^{2,4,15}. It is possible that both systems occur in cetaceans¹⁶, with some species employing one or the other method, depending on perceived frequencies. Both systems probably functioned in *Pakicetus*. The external ear bears strong resemblance to that of terrestrial mammals (no involvement of the lower jaw). The middle ear shows conditions consistent with inertial hearing (inflation of incus), but also shows conditions consistent with terrestrial mechanisms (crus longum longer than in other cetaceans). Thus, the hearing mechanism of *Pakicetus* is the only known intermediate between that of land mammals and aquatic cetaceans.

One specialization of the ear that occurs in all cetaceans, and in no other mammals, is the reorientation of the eardrum, malleus, and incus^{1,4}. The joint between the malleus and incus of most mammals consists of a saddle-shaped facet that faces rostromedially (right in Fig. 3). The same axis faces rostrolaterally (perpendicular to the plane of Fig. 3) in all described cetaceans. In *Pakicetus*, this facet faces more or less rostrally (Figs 2a and 3), implying that the position of the malleus is intermediate between those of a primitive terrestrial mammal and a cetacean. It has been stated that this rearrangement is caused by an evolutionary rotation^{4,17}, but no morphological intermediates were known until now. *Pakicetus* is the only

FIG. 2 Scanning electron micrographs of left incus of *Pakicetus* (H-GSP 91035). Approximate lateral (a), dorsocaudal (b), and mediostrual (c) views. Crus longum (C. lg.) and crus breve (C. br.) are least foreshortened in a. The incus was found in a poorly preserved tympanic that could be identified on the basis of the resemblance of its ventral side to that of the type specimen of *Pakicetus inachus* (where mainly the ventral side is visible)⁵. The resemblances include a deeply fissured groove for the auditory tube, a convex lateral side and a straight contact for the basioccipital.

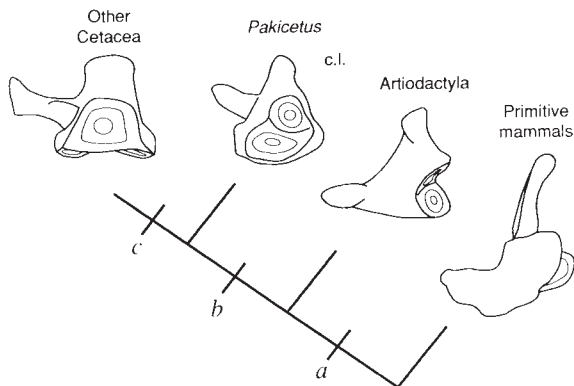
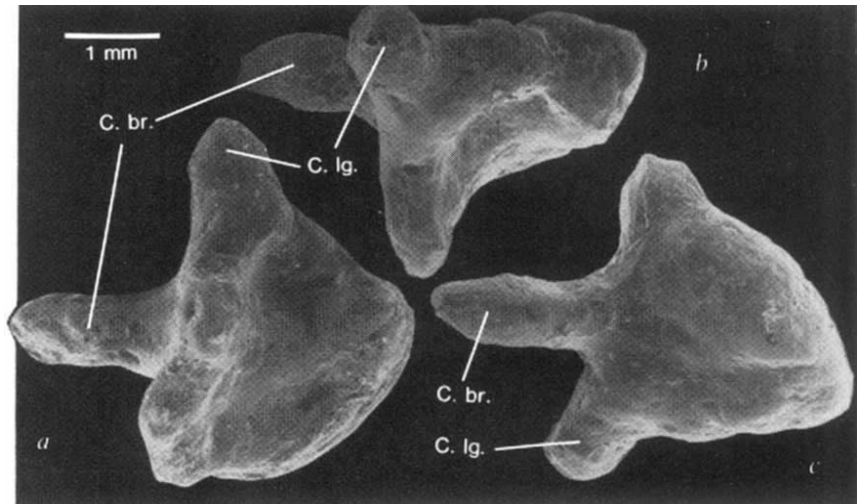


FIG. 3 Morphology of the incus and proposed phylogenetic relations of cetaceans. The *Pakicetus* incus is intermediate between those of modern artiodactyls and cetaceans in the moderately reduced length of the crus longum (c.l.), and the orientation of the facets for the malleus (thin contour lines). Defining characters are as follows: a, artiodactyls and cetaceans, crus longum shorter than crus breve; b, cetaceans, relatively inflated incudal body, shorter crus longum, and partly rotated malleal joint; c, cetaceans except *Pakicetus*, inflated crus longum and completely rotated malleal joint. Shown are the left incus of *Monodelphis* (primitive therian morphology), *Odocoileus* (artiodactyl), *Pakicetus* and *Ziphius* (modern cetacean). All incus are oriented at right angles to the plane of crura.

cetacean with partially rotated ear bones and thus substantiates the rotation hypothesis.

Cetaceans probably originated from extinct mesonychid condylarths¹⁸. Morphological⁸ and molecular data^{19,20} suggest that, among Recent mammals, cetaceans are more closely related to the ungulates, especially artiodactyls^{21,22}. The incus is rarely found in fossil mammals. No incus is known for any condylarth, but in Recent artiodactyls the bone is distinctive. The crus breve of artiodactyls is generally somewhat longer than the crus longum, whereas the reverse is true in most other mammals¹⁰. The *Pakicetus* incus has proportions similar to those of the artiodactyl incus, whereas the crus longum of all other cetaceans is further reduced in length and greatly inflated (Fig. 3). Crural length suggests that the incus of *Pakicetus* is more primitive than that of any other cetacean

and indicates that artiodactyls and cetaceans probably are closely related.

The morphology of the incus and jaw, as well as that of the bony connections between hearing organ and skull⁶, and the extreme shallowness of the river in which it was deposited suggest that *Pakicetus* was not fully aquatic, as has been previously suggested^{23,24}. A robust incus combined with a moderately long crus longum also occurs in some pinnipeds^{10,11}. True seals hear both in air and under water^{25,26}; their hearing may be based on a lever system in air, but on inertial lag under water^{27,28}. With respect to hearing, the ancestral cetacean seems to resemble modern true seals and walrus. This shows that the plethora of adaptations necessary for aquatic life were acquired stepwise, with initial retention of terrestrial adaptations. □

Received 30 October; accepted 7 December 1992.

- Fraser, F. C. & Purves, P. E. *Proc. R. Soc. Lond.* **152**, 62-78 (1960).
- Purves, P. E. in *Whales, Dolphins, and Porpoises* (ed. Norris, K. S.) 320-378 (Univ. California Press, Berkeley, 1966).
- Reysenbach de Haan, F. W. *Acta otolar. Suppl.* **134**, 1 (1957).
- Fleischer, G. *Adv. Anat. embryol. Cell Biol.* **55**, 1-70 (1978).
- Gingerich, P. D. & Russell, D. E. *Contr. Mus. Paleont. Univ. Mich.* **25**, 235-246 (1981).
- Gingerich, P. D., Wells, N. A., Russell, D. E. & Shah, S. M. I. *Science* **220**, 402-406 (1983).
- West, R. M. & Lukacs, J. R. *Contr. Biol. Geol. Milwaukee Publ. Mus.* **26**, 1-20 (1979).
- Slijper, E. J. *Whales* (Cornell Univ. Press, Ithaca, 1979).
- Norris, K. S. in *Animal Sonar Systems* (eds Busnel, R.-G. & Fish, J. F.) 495-509 (Plenum, New York and London, 1980).
- Doran, A. H. G. *Trans. Linn. Soc. Lond.* **1**, 371-497 (1879).
- Wyss, A. R. *Am. Mus. Novit.* **2871**, 1-31 (1987).
- Repenning, C. A. in *Functional Anatomy of Marine Mammals* (ed. Harrison, R. J.) 307-331 (Academic, New York, 1972).
- McCormick, J. G., Wever, E. G., Palin, J. & Ridgway, S. H. *J. acoust. Soc. Am.* **48**, 1418-1428 (1970).
- McCormick, J. G., Wever, E. G., Ridgway, S. H. & Palin, J. in *Animal Sonar Systems* (eds Busnel, R.-G. & Fish, J. F.) 449-467 (Plenum, New York and London, 1980).

- Purves, P. E. & Pilleri, G. E. *Echolocation in Whales and Dolphins* 1-261 (Academic, London, 1983).
- Ketten, D. R. in *The Biology of Hearing* (eds Webster, D., Fay, R. & Popper, A.) 717-750 (Springer, Berlin, 1991).
- Lancaster, W. C. *J. vert. Pal.* **10**, 117-127 (1990).
- Van Valen, L. *Bull. Am. Mus. Nat. Hist.* **132**, 10-126 (1966).
- Miyamoto, M. M. & Goodman, M. *Syst. Zool.* **35**, 230-240 (1986).
- Goodman, M., Czelusniak, J. & Beebe, J. E. *Cladistics* **1**, 171-185 (1985).
- Novacek, M. J. *Nature* **356**, 121-125 (1992).
- Wyss, A. *Nature* **347**, 428-429 (1990).
- Oelschläger, H. A. *Am. J. Anat.* **177**, 353-368 (1986).
- Oelschläger, H. A. *Gegenbaurs Morph. Jahrb.* **133**, 673-685 (1987).
- Watkins, W. A. & Wartzok, D. *Mar. mam. Sci.* **1**, 219-260 (1985).
- Moore, P. W. B. & Schusterman, R. J. *Mar. mam. Sci.* **3**, 31-53 (1987).
- Ramprasad, F., Corey, S. & Ronald, K. *Can. J. Zool.* **49**, 241-248 (1971).
- Ramprasad, F., Corey, S. & Ronald, K. *Can. J. Zool.* **51**, 589-600 (1973).

ACKNOWLEDGEMENTS. We thank D. P. Domning, P. D. Gingerich, F. Husain, W. L. Hylander, D. R. Ketten, M. Raza, D. W. Smith, K. K. Smith, M. Arif and C. F. Ross; M. C. Maas provided the scanning photographs, and the National Geographic Society funded field work.

Selection of exaggerated male traits by female aesthetic senses

Magnus Enquist & Anthony Arak

Department of Zoology, University of Stockholm,
S-10691 Stockholm, Sweden

DARWIN¹ suggested that many apparently deleterious secondary sexual characters in males, such as bright colours, elaborate ornaments and conspicuous displays, evolved as a result of female choice. Darwin never tried to explain the crucial agent of selection, that females have preferences for exaggerated male traits. Rather, he took it for granted that females of many species possess a 'sense of the beautiful', akin to the aesthetic sense in humans. The question of why such preferences evolve remains a controversial issue^{2,3}. Here we report that mechanisms concerned with signal recognition possess inevitable biases in response that act as important agents of selection on signal form. The existence of such biases may be sufficient to explain the evolution of exaggerated male secondary sexual traits, and elaborate signals in general.

To understand the evolution of exaggerated traits and conspicuous displays used by males to attract females it is first necessary to explain why females prefer such traits. In line with modern thinking in evolutionary biology, recent attempts to explain female choice focus on selection and adaptive behaviour. The problem of how females recognize conspecific males is often ignored because recognition is usually assumed to be perfect and without biases.

All but the simplest of recognition systems are unlikely to be perfect. Taking the visual system as an example, perfect recognition implies that an animal unerringly reacts to all images of the correct object (or class of objects) and never reacts to all other, inappropriate images. But a recognition mechanism can only be expected to react appropriately to those images it has been selected to identify. One cannot predict with certainty how an animal will react to new images it experiences. Many will have no effect, but because there is an almost infinite number of possible images that the retina may experience, it is expected that some of these will elicit a greater response than the particular signals to which the system has been selected to respond.

Of particular interest are stimuli that are similar to those the

animal has been selected (or trained) to respond to because they may be affected by generalization. Generalization is a common property of recognition systems that involves classifying novel variations of stimuli into particular categories that the organism has experienced before. Such a property is necessary because variation in orientation, distance, light conditions and backgrounds can give rise to millions of possible images of a given object on the retina. A recognition system cannot be pre-programmed or trained to identify each such image separately.

To explore the role of the recognition in the evolution of signals, we have studied some examples of recognition mechanisms based on artificial neural networks^{4,5}. Even the most simple artificial networks, consisting of a few interconnected cells, exhibit many of the properties shown by animal recognition systems: they are easily trained to classify objects and perform generalizations. Such networks provide convenient tools for uncovering general principles of recognition free from much of the complexity found in the nervous systems of real organisms.

A problem was first presented to a network representing the recognition system of a female bird (see Fig. 1 for details). The problem was for the female to recognize males of her own species in the presence of a similar sympatric species using simple visual cues. Males of the two species were identical in all respects except that conspecific males had slightly longer tails than heterospecifics.

The network was 'trained' by a procedure that mimics the process of natural selection of recognition systems that occurs over evolutionary time. The network quickly evolved the ability to distinguish images of conspecific males from heterospecific males and random patterns. Moreover, there was no single evolutionarily stable mechanism of recognition; instead it was found that many different networks, that is, combinations of connection weights, could achieve the recognition task with virtually no errors (only two examples are illustrated in Fig. 2). When the networks' reactions to new stimuli were investigated it was apparent that most stimulus patterns were not effective in eliciting a courtship response. But some stimuli were super-normal⁶ (they elicited stronger responses than any of the training stimuli). Some of these stimuli bore no resemblance to the training stimuli (Fig. 2e), whereas others (Fig. 2a, b, c) appeared to be exaggerated forms of the conspecific male (longer tails, more tails and so on), illustrating the phenomenon of peak

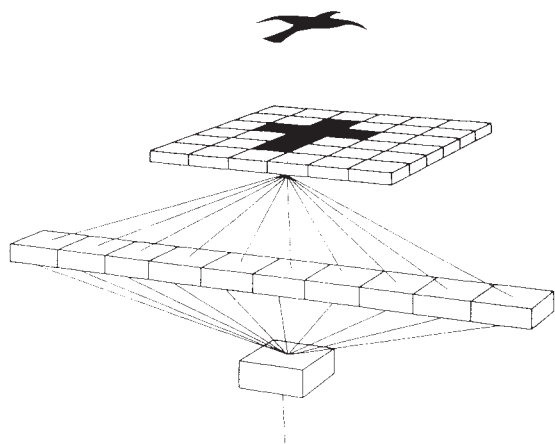


FIG. 1 A simple artificial neural network is shown which consists of a retina 6 × 6 receptor cells, 10 hidden cells and one output cell. Each cell in one layer connects to all cells in the next layer and to each connection a weight is associated that regulates the strength of the signal passing between cells (For clarity, only those connections originating from one of the receptor cells are illustrated). When the network is stimulated each of the receptor cells receives an input of zero or one. The output from these cells equals the input. The input to all other cells, that is cells in the hidden layer and the output cell, is a weighted sum of the output from all cells in the previous layer. The output from a hidden cell or an output cell is a sigmoid function of its net input. The network was said to recognize a subset of patterns if these patterns gave rise to an activity (output) in the output cell that was greater than a certain threshold, whereas all other patterns that occur give rise to activities below this threshold. By presenting patterns to the retina, the network can evolve the ability to discriminate between two or more groups of images. Starting with some (random) vector of connection weights, a new network was first created by mutating some of the weights. The probability of mutation for a particular connection weight was 0.1 and when a mutation occurred, an increment drawn from a normal distribution ($\sigma = 0.1-0.4$) was added to the weight. The performance of the new network in the recognition task was then compared with the original one and the best retained. This iteration continued until the probability of an incorrect decision by the network was less than 10^{-5} . A courtship reaction to an image was assumed to occur when the sum of external (stimulus) and internal (motivational factors) reached the threshold value of 0.5. The internal factor was assumed to vary independently of the external factor and to be normally distributed with $\mu = 0$ and $\sigma = 0.02$.

Recognition task				Test images				
Reject		Accept		(a)	(b)	(c)	(d)	(e)
Random								
N:	20	16	48	32	16	24	60	16
I:	0.26	0.42	0.60	0.63 (66%)	0.68 (91%)	0.76 (100%)	0.34 (0%)	0.59 (50%)
II:	0.26	0.41	0.61	0.65 (74%)	0.68 (88%)	0.78 (100%)	0.35 (0%)	0.63 (57%)

FIG. 3 Results of a simulation in which both male tail length and the female recognition system coevolve. It was assumed that selection favours those males that elicit the strongest courtship response in females, and those females that make the fewest mistakes in choosing the correct species. In the first step of the simulation, the trained network was presented with images of 'mutant' conspecific males with various tail lengths, and the first image that elicited the strongest courtship response was selected. In the second step, the network was allowed to mutate until it discriminated between the selected males and heterospecifics with fewer errors than before. The evolutionary process was studied by repeated iteration of this procedure. The image eliciting the highest response after 100 iterations depends on the survival cost associated with increases in tail size. Longer tails evolve when there is only a weak effect of tail size on survival. As male tail length increases, females become less efficient at recognizing males with the original tail size as conspecifics. * Survival of selected form after 100 steps of simulation. Fitness = $P \times$ mating success.

shift⁷. Images of males without tails (Fig. 2 *d*) gave a weaker response than males with short tails (heterospecifics), suggesting that generalization occurs in both directions. This proves that the recognition mechanism itself exerts selection pressure on the signal; in this case, there is a 'bias' in the mechanism favouring males with longer tails.

We investigated the consequences of this bias in an evolutionary simulation by allowing both male tail length and female recognition system to change by mutation. In general, the simulation resulted in the evolution of longer tails in males, alongside a decrease in female responsiveness to conspecific males with the original tail length (Fig. 3). Exaggeration occurs even in cases when increasing tail size decreases the survival of males, but the extent of exaggeration is inversely related to survival cost. These results confirm Darwin's view¹ that traits that give an advantage in mating can evolve to such extremes that they decrease male survival. The model also demonstrates that female preference for extreme male traits can evolve simply as a consequence of the need for females to recognize males of their own species, not to discriminate between them.

It is well known that sense organs often show biases in their response to signals along certain dimensions. It follows that such biases could act as important agents of selection on the form of signals. Indeed, a few recent studies⁸⁻¹¹ present convincing evidence that biases in the sensory apparatus of females towards certain signals may have existed before the appearance of the same signals among males. But the question of why such biases exist has been largely ignored. Our models suggest that biases in response to signals inevitably exist as a fundamental consequence of the context in which recognition occurs. Because the number of forms that a signal can take is almost infinite,

FIG. 2 The network was trained to discriminate long-tailed birds (conspecifics) from short-tailed birds (heterospecifics) and random images. The images were projected onto the retinal array in numerous positions and rotations, the total number of projections of each image type denoted by N . The average response to each image type after training is shown below each image for two examples of different networks (I and II) each capable of solving the recognition task. The threshold value for a courtship response is 0.5. After training, novel 'test' images are projected onto the grid and the response of the network measured. The network shows the strongest response to bird-like images with even longer tails (*a*, *b*) or longer wings (*c*) than it has been trained to recognize, but it does not react to images of birds without tails (*d*). Some fairly abstract patterns (such as *e*) also elicit a supernormal response. The figures in parentheses show the percentage of presentations of each image type that gave rise to a higher response than the image of the conspecific male.

Evolutionary simulation						
Tail size	(a)	(b)	(c)	(d)	(e)	
Probability	2	0.50*	0.50	0.50	0.50	0.50
of survival (P)	3	0.25	0.30*	0.35*	0.40	0.45
	4	0	0.10	0.20	0.30*	0.40*
Resulting phenotype:						
Recognition of original (Tail size = 2)	100%	88%	84%	58%	59%	

the recognition mechanism is always likely to show a greater response to some variants of the male signal not yet in existence. Such unexpressed, or 'hidden', female preferences will change continuously as a side effect of selection for improved recognition, by genetic drift (many solutions exist for a given recognition problem) and because of correlated effects¹² of selection acting on male signals. Moreover, because biases may occur simultaneously along several dimensions of the signal, selection on the recognition mechanism in one dimension may result in many dimensions of bias. Because female preferences and male traits used in signalling are very unlikely to be at equilibrium they can be expected to appear and disappear at a greater rate than would be the case for other aspects of morphology and behaviour. Although the form of male courtship signals will typically evolve away from all other stimuli that the female regularly experiences, the precise course that evolution takes will be highly unpredictable. This instability may explain the great diversity in the form of signals observed in courtship behaviour, even within single species.

Finally, the process of exaggeration described here is not confined to signals used for mate attraction; it applies with equal force to all contexts of signalling, including interspecific communication (such as warning colouration), and may offer a general explanation for the elaboration of signals that occurs during the process of ritualization¹³. It is an interesting thought that all nervous systems built for recognition may share certain general biases which result from hidden properties of the recognition system. Indeed, many elaborate signals that occur in nature are often as impressive to human observers as they appear to be to the intended recipient. Darwin's idea that a 'sense of the beautiful' is an inherent, aesthetic property of animal nervous

systems may be not far from the truth. In Darwin's own words 'When we behold a male bird elaborately displaying his graceful plumes or splendid colours . . . it is impossible to doubt that [the female] admires the beauty of her male partner'¹. □

Received 8 October; accepted 19 November 1992.

1. Darwin, C. *The Descent of Man and Selection in Relation to Sex* (Murray, London, 1871).
2. Kirkpatrick, M. & Ryan, M. J. *Nature* **350**, 33–38 (1991).
3. Maynard Smith, J. *Trends Ecol. Evol.* **6**, 146–151 (1991).
4. Caudill, M. & Butler, C. *Naturally Intelligent Systems* (MIT Press, Cambridge, 1990).
5. Eberhart, R. C. & Dobbins, R. W. *Neural Network PC Tools* (Academic, San Diego, 1990).
6. Tinbergen, N. *Wilson Bull.* **60**, 6–52 (1948).
7. Spence, K. W. *Psychol. Rev.* **44**, 430–444 (1937).
8. Basolo, A. L. *Science* **250**, 808–810 (1990).
9. Basolo, A. L. *Science* **253**, 1426–1427 (1991).
10. Ryan, M. J., Fox, R. C., Wilczynski, W. & Rand, A. S. *Nature* **343**, 66–67 (1990).
11. Ryan, M. J. *Oxf. Surv. evol. Biol.* **7**, 156–195 (1991).
12. Lande, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3721–3725 (1981).
13. Huxley, J. S. *Phil. Trans. R. Soc. B251*, 249–271 (1966).

ACKNOWLEDGEMENTS. We thank N. Davies, O. Leimar and R. Rosenberg for their comments on an earlier version of the manuscript.

Involvement of an orthologue of the *Drosophila* pair-rule gene *hairy* in segment formation of the short germ-band embryo of *Tribolium* (Coleoptera)

Ralf J. Sommer* & Diethard Tautz

Zoologisches Institut der Universität München, Luisenstrasse 14, D-8000 München 2, Germany

* Present address: California Institute of Technology, Division of Biology, Pasadena, California 91125, USA

THE segments in long germ-band insect embryos, like *Drosophila*, are all determined at syncytial blastoderm stage. This is in contrast to short germ-band embryos which show an early determination of only the anterior head segments, whereas the more posterior thoracic and abdominal segments are sequentially added after formation of a primary germ anlage (reviewed in ref. 1). Segment formation in *Drosophila* involves the pair-rule genes which define double segmental periodicities^{2,3} and which have been considered to represent a special adaptation to the long germ-band type development^{4,5}. *hairy* belongs to the primary pair-rule genes in *Drosophila* which are directly regulated by the gap genes, such as

Krüppel^{6–13}. We have isolated the orthologues of *hairy* and *Krüppel* from the flour beetle *Tribolium castaneum* which has a short germ type development¹⁴. We show here that *hairy* is expressed in several stripes at blastoderm stage and later on in two stripes in the growth zone of the developing embryo. *Krüppel* expression overlaps *hairy* stripe three and four expression, very similar to *Drosophila*. This suggests that the segment patterning mechanism that acts in an open blastoderm in *Drosophila* works in a similar way in the cellularized *Tribolium* embryo.

To understand the potential role of pair-rule genes in short germ-band embryos, we cloned the orthologues of *Krüppel* and *hairy* from the flour beetle *Tribolium*. *Krüppel* codes for a zinc-finger protein of the Cys₂-His₂ type¹⁵. We have previously developed rules for the generation of polymerase chain reaction (PCR) primers to clone these type of genes¹⁶, allowing us to obtain *Krüppel* orthologues from a variety of arthropods, including *Tribolium*¹⁶. *hairy* codes for a transcription factor with a helix-loop-helix motif¹⁷ allowing the design of PCR primers within conserved portions of this domain (Fig. 1). An orthologous *hairy* fragment was obtained in this way and used to clone the respective genomic region from a library¹⁸ (Fig. 1).

Embryogenesis in *Tribolium* starts with the formation of a syncytial blastoderm (Fig. 2a, b), similar to *Drosophila*. But instead of proceeding directly with the formation of the segments, as would be typical for long germ-band embryos like *Drosophila*, it forms first a germ anlage (Fig. 2c, d). This germ anlage includes the gnathal segments and a growth zone from which the remaining segments are formed (see legend to Fig. 2e, f).

The expression pattern of *Krüppel* and *hairy* in *Tribolium* was analysed by whole-mount *in situ* hybridization¹⁹. The first expression of *hairy* is seen at early blastoderm stage (Fig. 3a) in two circumferential stripes. These stripes move towards posterior (Fig. 3b) and lose their dorsal expression, but remain ventrally expressed in the region of the germ rudiment (Fig. 3c). A new expression domain of *hairy*, formally the third stripe, forms now at the posterior end of the embryo (Fig. 3b, c). In *Drosophila*, the third and the fourth *hairy* stripes lie within the region of the *Krüppel* expression domain^{10,11}. Analysis of *Krüppel* expression at this stage indicates that the same is true for *Tribolium*. The first expression of *Krüppel* is seen at the posterior end, preceding the expression of *hairy* in the same region (Fig. 3d). As shown in Fig. 2, the amnion folds in this region and grows towards the anterior. At the same time, cell proliferation occurs at the posterior end of the germ rudiment. The effect of these separate growth and cell movement processes is that the cells which were at the posterior end at blastoderm stage will become part of the central region of the early germ band, where

FIG. 1 Protein sequence comparison (single-letter amino-acid code) between the *hairy* orthologues from *Drosophila* and *Tribolium*. Dashes denote identities, dots denote insertions/deletions. The positions of the introns in both species are denoted by arrowheads, the helix regions of the helix-loop-helix motif are underlined and the arrows indicate the positions of the PCR primers. The *Drosophila* sequence (numbering according to ref. 17) starts with codon 33 (position 1,611), directly after the first intron. The *Tribolium* sequence is derived from a genomic clone and shows an intron at the same position, the first exon of *hairy* in *Tribolium* could not yet be identified. *Drosophila* shows another intron after codon 65 which is also present in *Tribolium* (531 base pairs long). The sequence comparison shows that not only the helix-loop-helix motif is conserved, but also several additional regions within the coding region, suggesting that a true orthologue was obtained. There is, in contrast, a clear divergence in the regions of high cryptic simplicity²⁴, for example the glutamine and alanine stretches, which has similarly been noted in interspecific comparisons between other genes^{24,25}.

<i>Drosophila</i>	SNKPMIEKRRRARINNCINELKTLITLDATKKDPARHSKLEKADILEKTVKHLQETQRQQAAMQQAADPKIVHKFKAGFAD
<i>Tribolium</i>	-----S-----M-----N-----M-----W-PT--SV-S--R--SE
<i>Drosophila</i>	CVNEVSFRFPGIEPAQRRLRLQHLSHCIN.GVKTELHQQRQQQQSIHAQMI.PSPSSPEQDSQQGAAPYIFGIQQTAS
<i>Tribolium</i>	-AS--G-----LD-VVR-----AS-L-Q-Q-EPQV-----VTVPEVAFNNI
<i>Drosophila</i>	GYFLPNGMQVPTKLPNGSIALVLPQSLPQQQQQQQLLHQHQQQQQLAVAAAAAQAQQPMLVSMFQRTASTGSSASH
<i>Tribolium</i>	ILNGT-V-LV-IR-A--D-----TQGAS-----PL-L--PI-----A
<i>Drosophila</i>	SSAGYESAPGSSSSCSYAPP.SPANSSYEPMDIKPSVIRVPMQQLSLVKKQI.KEEEPKRW
<i>Tribolium</i>	-----N-S-SQ--E-----ESVR-----VRRREPT--R-----VVEIVM

METHODS. A fragment of *hairy* from *Tribolium* was amplified by PCR²⁶ using the following primers CCTATATGAGAA(A,G)(C,A,G)(A,C,G,T)(A,C)G and ATGCTTGAC(A,G)GTCTT(C,T)TC(A,C,G,T)A (corresponding to positions 1,621–1,640 and 1,905–1,886 in the *Drosophila* sequence, ref. 17) and cloned into an M13 vector as described¹⁶. The fragment was then used to screen a genomic library of *Tribolium*¹⁸ and the sequence of the coding region was obtained. The fragment used for the *in situ* hybridization experiments included the whole region shown, apart of the first four and the last 20 amino acids.

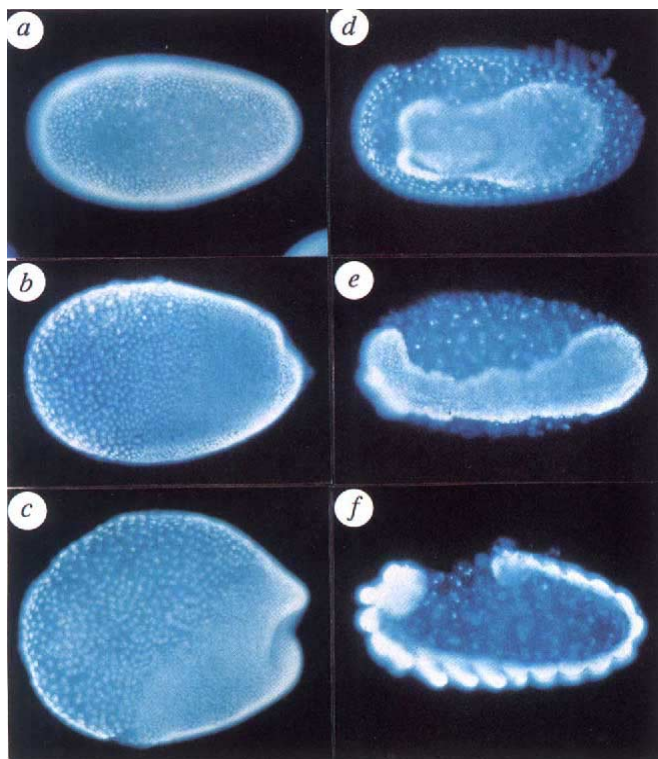


FIG. 2 Early development of *Tribolium* embryos. The figure shows progressively older embryos stained with DAPI (nuclear staining). All embryos (apart of the one in *d*) are oriented anterior to the left and dorsal up. The size of the *Tribolium* egg is about 0.6×0.4 mm. *a*, Blastoderm stage with homogeneous distribution of the nuclei around the egg cortex. *b*, Enrichment of nuclei at the posterior ventral side, caused by a movement of some of the more anterior nuclei towards posterior positions¹ and by further nuclear divisions in the respective region. *c*, Completion of germ rudiment formation and beginning of the folding of the amnion. The amnion grows then towards anterior and forms eventually a cell layer ventrally beneath the germ band. *d*, Early germ anlage surrounded by the cells of the extra embryonic serosa, viewed from the ventral side; amnion formation is almost complete, apart of an oval area in the anterior region of the germ band. *e*, Embryo during segmental growth; cell division takes place predominantly in the posterior growth zone, but also at more anterior positions. *f*, Embryo after the completion of the segmentation process. The early embryonic development of a closely related species (*Tribolium confusum*) is described in more detail in¹⁴. Note that the older literature would classify *Tribolium* as an intermediate rather than a short germ type embryo. A re-evaluation of the criteria for this classification suggests however that this distinction is only quantitative rather than qualitative¹. We use therefore 'short germ embryo' as a generic term of all types of insect embryos that show a segmental growth process after blastoderm stage.

they are involved in the formation of the thoracic segments and anterior abdominal segments. *Krüppel* expression persists in the respective cells, forming a single broad expression domain (Fig. 3*e, f*). In contrast, the *hairy* domain is now split into two stripes (Fig. 3*g, h*) but still lies within the realms of the *Krüppel* domain (compare Fig. 3*e, f* with *g, h*). Note that *hairy* stripes three and four form at first a single domain in *Drosophila* as well, which later splits into the two stripes⁶⁻¹¹.

Thus, four *hairy* stripes have formed until this stage of development in *Tribolium*, whereby expression in *hairy* stripes 1 and 2 has ceased already, at least at the RNA level. We note that the formation of the *hairy* stripes occurs in an anterior-posterior order. This is in contrast to the morphological differentiation of the segments in insects, which starts in a region called the differentiation centre²⁰ and which is normally located in the region of the prospective thoracic segments²¹. This indicates that the timing of the molecular differentiation is not necessarily coupled with the timing of the morphological differentiation (compare also ref. 1).

Our results indicate that the spatial relationships of the expression domains of *hairy* and *Krüppel* as well as their relative locations with respect to the developing segments are conserved between *Drosophila* and *Tribolium*. We suggest therefore that the two genes may have a truly homologous function and in particular that *hairy* has a pair-rule function in *Tribolium* at least at these early stages in development.

The largest difference between long and short germ type development is the difference in the generation of the posterior segments. These are determined in an open blastoderm in long germ embryos, whereas they proliferate from a cellularized growth zone in short germ embryos. This would suggest that the molecular mechanism forming these segments might be different for the two types of development⁵. We have therefore analysed the later embryonal stages for signs of *hairy* expression. Figure 3*i* shows an embryo which is half-way through the segmentation process, with a clearly demarcated growth zone

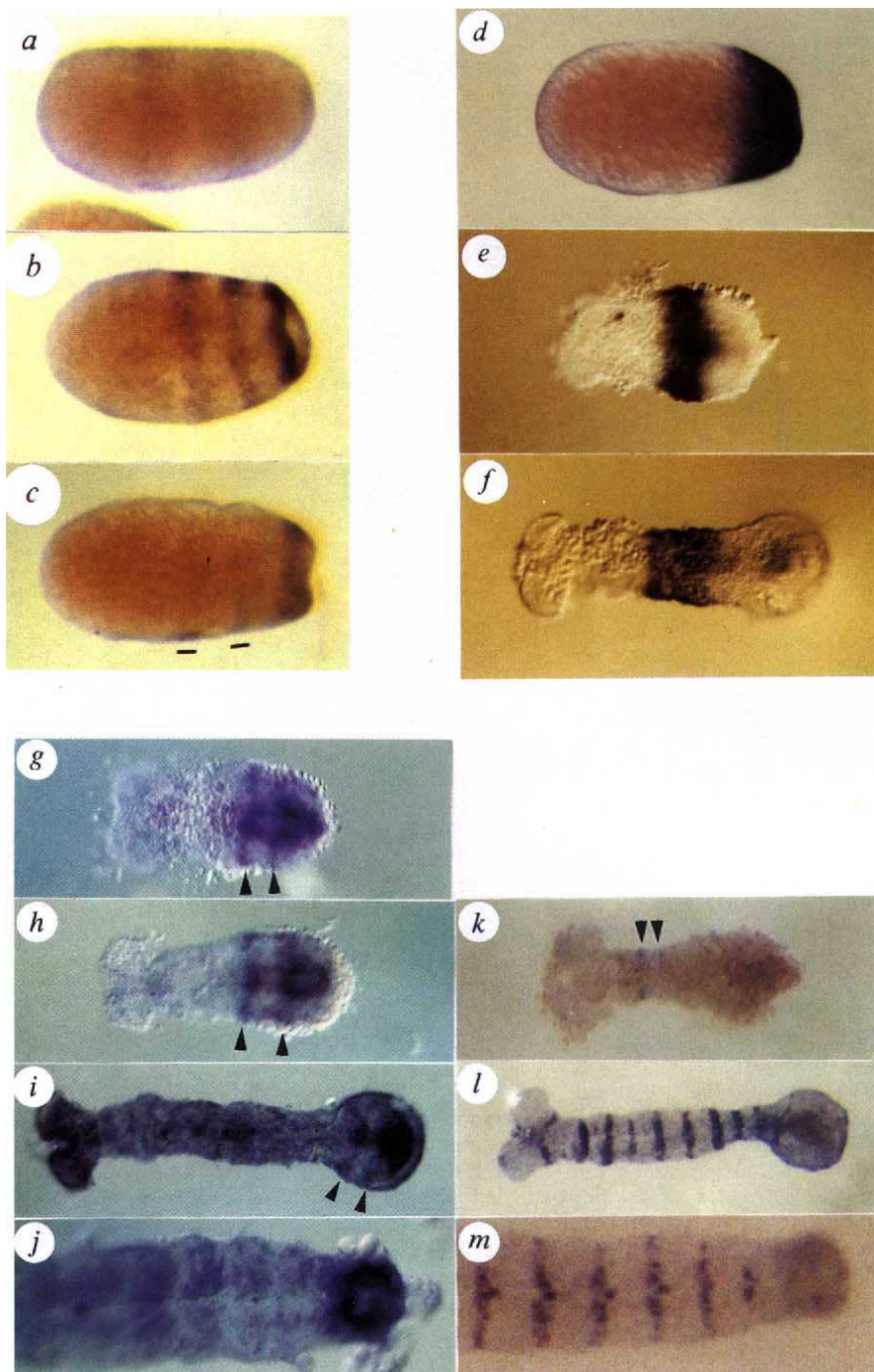
at the posterior end. *hairy* has two separate expression patterns in this posterior growth zone, a non-segmental expression in the mesoderm, which is seen throughout development (compare Fig. 3*g-i*), and two stripes in the ectoderm (marked with arrowheads). The latter expression indicates that *hairy* is probably involved in segment determination even at this stage. To analyse this in more detail, we have also stained the embryos with a monoclonal antibody against the *engrailed* gene product²². *engrailed* belongs to the class of segment polarity genes which comprise the last step in the molecular hierarchy of segment formation and which are regulated by the pair-rule genes³. *hairy* expression must therefore precede *engrailed* expression, if it is causally involved in the segmentation process. We find that this is indeed the case. The first signs of *engrailed* expression become visible only well after the *hairy* stripes are established (compare Fig. 3*h, k*). Most importantly though, there is no *engrailed* staining in the growth zone of the embryo (Fig. 3*l*), indicating that the patterning mechanism leading to the formation of the *hairy* stripes precedes the formation of the segmental borders.

It is unclear how this patterning in the growth zone is achieved. A mechanism homologous to that found in *Drosophila*, with a direct action of morphogenetic transcription factors²³, would seem unlikely, given that the growth zone is apparently fully cellularized. In contrast, it cannot be excluded that the cells in this region communicate through special junctions which would allow the direct exchange of transcription factors.

Our data provide the first evidence for a pair-rule gene being involved in the segmentation of short germ-band embryos. A similar analysis with the pair-rule gene *even-skipped* in *Schistocerca* has indicated that this gene is not involved in the segmentation process, but is only expressed in the nervous system⁴. It seems therefore possible that not all pair-rule genes are evolutionary conserved. In contrast, the striking parallels in *hairy* expression between *Drosophila* and *Tribolium* shown here suggest that the patterning mechanism through a transient

FIG. 3 Expression of *hairy* (a–c, g–j), *Krüppel* (d–f) and *engrailed* (k–m) in *Tribolium* embryos. The figure shows whole-mount *in situ* hybridizations for *hairy* and *Krüppel*, as well as antibody staining for *engrailed*. The embryos are oriented with the anterior to the left. a–d, Blastoderm stages, e–m, embryos prepared free from the surrounding egg material. a, Earliest *hairy* staining in two circumferential stripes. b, c, *hairy* staining at the posterior pit stage showing that the earlier stripes have shifted towards posterior; b, ventral view and c, lateral view, showing that the stripes are present only ventrally (indicated by bars), but not dorsally. d, *Krüppel* staining at blastoderm stage. The embryo shown was counterstained with DAPI and found to be at early blastoderm stage (comparable to Fig. 2a). e, f, *Krüppel* expression in the early germ anlage at the beginning of the segmental growth process. *Krüppel* expression fades at later stages, but is again expressed in the proctodaeum after the completion of segment formation, similar to as is shown for *hairy* in j, g, *hairy* expression in an early germ anlage showing two stripes, as well as expression of *hairy* in the mesoderm. h, k, Embryos at the beginning of the segmental growth process showing *hairy* staining in two posterior stripes (h) and the first *engrailed* staining (k) in anterior stripes (marked with arrowheads). j, l, Embryos in the middle of the segmental growth process showing *hairy* stripes in the growth zone (j) (labelled with arrowheads) and *engrailed* stripes anterior to the growth zone (l). j, m, Posterior region of embryos at the end of the segmental growth process. *hairy* staining is evident around the proctodaeum (j), whereas *engrailed* demarcates the posterior segmental boundaries (m).

METHODS. The flour beetle *Tribolium castaneum* was grown at 25 °C on white flour supplemented with yeast. Eggs were collected as described²⁷. Dechoriation, fixation and whole-mount *in situ* hybridizations were basically done as described for *Drosophila*¹⁹. Vitellin membranes were removed by repeatedly passing the eggs through a hypodermic needle (1.1 mm diameter) after fixation. The hybridization probes were single-strand labelled by PCR²⁸ with digoxigenin (Boehringer-Mannheim).



pair-rule pre-pattern is a fundamental component of the segmentation process in insects. □

Received 3 September; accepted 24 November 1992.

- Sander, K. *Adv. Insect Physiol.* **12**, 125–238 (1976).
- Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795–801 (1980).
- Ingham, P. *Nature* **335**, 25–34 (1988).
- Patel, N., Ball, E. & Goodman, C. *Nature* **357**, 339–342 (1992).
- Akam, M. & Dawes, R. *Curr. Biol.* **2**, 395–398 (1992).
- Ingham, P. W., Howard, K. R. & Ish-Horowitz, D. *Nature* **318**, 439–445 (1985).
- Ingham, P. W., Ish-Horowitz, D. & Howard, K. *EMBO J.* **5**, 1659–1665 (1986).
- Hooper, K. L., Parkhurst, S. M. & Ish-Horowitz, D. *Development* **107**, 489–504 (1989).
- Howard, K. & Struhl, G. *Development* **110**, 1223–1231 (1990).
- Pankratz, M. *et al. Cell* **61**, 309–317 (1990).
- Riddihough, G. & Ish-Horowitz, D. *Genes Dev.* **5**, 840–854 (1991).
- Preiss, A., Rosenberg, U. B., Kienlin, A., Seifert, E. & Jäckle, H. *Nature* **313**, 27–32 (1985).
- Gaul, U., Seifert, E., Schuh, R. & Jäckle, H. *Cell* **50**, 639–647 (1987).
- Stanley, M. S. M. & Grundmann, A. W. *Ann. ent. Soc. Am.* **63**, 1248–1256 (1970).
- Rosenberg, U. B. *et al. Nature* **319**, 336–339 (1986).

- Sommer, R. J., Retzlaff, M., Goerlich, K., Sander, K. & Tautz, D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10782–10786 (1992).
- Rushlow, C. *et al. EMBO J.* **8**, 3095–3103 (1989).
- Brown, S. J., Henry, J. K., Black IV, W. C. & Denell, R. E. *Insect. Biochem.* **20**, 185–193 (1990).
- Tautz, D. & Pfeifle, C. *Chromosoma* **98**, 81–85 (1989).
- Krause, G. *Biol. Zbl.* **59**, 495–536 (1939).
- Counce, S. J. *A. Rev. Entomol.* **6**, 295–312 (1961).
- Patel, N. H. *et al., Cell* **58**, 955–968 (1989).
- Pankratz, M. J. & Jäckle, H. *Trends Genet.* **6**, 287–292 (1990).
- Tautz, D., Trick, M. & Dover, G. *Nature* **322**, 652–656 (1986).
- Treier, M., Pfeifle, C. & Tautz, D. *EMBO J.* **8**, 1517–1525 (1989).
- Saiki, R. K. *et al. Science* **239**, 487–491 (1988).
- Sokoloff, A. *The Biology of Tribolium* (Clarendon, Oxford, 1972).
- Patel, N. & Goodman, C. S. in *Non-radioactive Labelling and Detection of Biomolecules* (ed. Kessler, C.) 377–381 (Springer, Heidelberg, 1992).

ACKNOWLEDGEMENTS. We thank I. Rambold and N. Folkman for technical assistance, S. Brown for providing the *Tribolium* library, N. Patel for the *engrailed* antibody, K. Sander, R. Schröder, M. Retzlaff, M. Hülkamp and M. Friedrich for discussions and helpful suggestions. R.J.S. was supported by a fellowship of the Verband der chemischen Industrie. The work was supported by the DFG (Gerhard Hess program to D.T.).

A fate map for the first cleavages of the zebrafish

David Strehlow* & Walter Gilbert†

* Boston University Hospital, Department of Biomolecular Medicine, 88 East Newton Street, Boston, Massachusetts 02118-2393, USA

† Harvard University, Department of Cellular and Developmental Biology, Biological Laboratories, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

A FATE map depicts how lineages from a defined blastomere will be restricted to particular tissues. In the amphibian *Xenopus laevis*, the early blastomeres of the embryo have traceable lineages¹⁻⁴. However, it has been thought that mixing of the blastomeres of *Brachydanio rerio* prevents them from giving rise to a fate map⁵⁻¹⁰. For this reason the zebrafish has been used as an example of a class of embryos, including mammals, which apparently exhibit little fate until later in embryogenesis. The amphibia were the exception. We report here the use of fluorescent dyes on very-high-molecular-mass carriers to trace a fate map from the first three cleavages of the zebrafish embryo to the three principal adult body axes: the first cleavage defines dorsal-ventral; the second cleavage, left-right; and the third, anterior-posterior.

Kimmel and Law¹¹ showed that lineage tracing of cleavage-stage zebrafish blastomeres was not feasible with the 17K (M_r 17,000) fluorescent dyes they used. Molecules of that size leaked from one blastomere to another and into the yolk. To circumvent these difficulties, we have used the simultaneous injection of two very-high-molecular-mass fluorescent dye/dextran polymers into different blastomeres at the two, four and eight-cell stages. The large size of the dyes, 2,000K (synthesized according to ref. 12), prevents leakage between the blastomeres or out of

blastomeres into the yolk at all stages, even as early as the two-cell stage (Fig. 1). The simultaneous use of two differently coloured dyes clarifies the differences in lineages, because both dyes can be visualized concurrently in the same developed animal. Figure 1 shows that the dyes are faithfully inherited by all cells in the lineage of the injected cell.

These techniques reveal that the three axes of the adult zebrafish, dorsal-ventral, left-right and anterior-posterior, are defined by lineage progenitors from the earliest stages. Figure 2a summarizes the fates that we attribute to each of the eight blastomeres.

Figure 3 shows examples of pairwise injections of eight-cell stage blastomeres with two different coloured dyes, using instances of blastomeres injected at both extremes of an axis to stress the differences in tissue specification along that axis. We traced a total of 92 blastomere lineages from the eight-cell stage. The photographs in A, B and C of the figure emphasize the dorsal-ventral (DV), left-right (LR), and anterior-posterior (AP) axes, respectively.

Figure 3A shows the dorsal-ventral pattern (DV): the dorsal pattern consists primarily of notochord (arrowhead in Ac), floor plate (arrow in Ae), dorsal epithelium, including the dorsal fin (arrowhead in Ae), and the hatching gland (arrowhead in Ad). The ventral pattern consists primarily of muscle mesoderm (arrow in Ab), ventral epithelium (arrowheads in Cb), the primordial nephric tubules (right arrowhead in Ab), the heart (left arrowhead in Ab) and blood. Labelling of the brain varies with the anterior-posterior axis as described below. Labelling

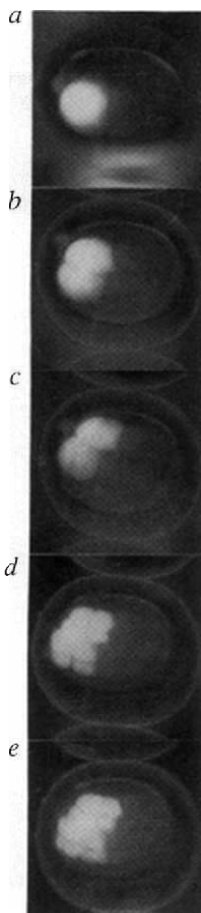


FIG. 1 The depicted embryo was injected in one cell at the two-cell stage with about 0.3 nl of a 25 mg ml⁻¹ solution of 2,000K Texas-red-conjugated dextran. a, Embryo immediately after injection. The non-fluorescent cell is seen in faint outline because it was photographed with a small amount of white light. b-e, Same embryo in the 4-, 8-, 16- and 32-cell stages, respectively. Each picture was taken ~30 min after the preceding one. At each point, half of the cells carry the fluorescent dye. This embryo was raised at room temperature.

METHODS. Fluorescein- or Texas-red conjugated dextrans (2,000K) were each injected into separate blastomeres in various positions. Positions not indicated in Fig. 3 were also injected. For the 8-cell stage fate map described, a total of 92 blastomeres were examined from 46 scored embryos. In general, 37% of the injected embryos were scored. We examined each embryo 5 min, 4 h and 22 h after the injection and rejected about 38% of the embryos as 'badly injected'. Another 25% of the embryos did not survive overnight after the injection or were rejected because of presence of fluorescence not confined to cells (presumably reflecting death of injected cell lineages). Animals were raised at 28 °C for 26 h after injection; then fluorescence patterns were observed and recorded both on video tape and 35 mm photographic film. Live embryos were anaesthetized and mounted between two 24 × 60 mm coverslips and the red and green fluorescence of both sides of the embryos scored. For dorsal observation of the left-right axis, animals were anaesthetized and immobilized in a mould of 1% agar in the end of a pasteur pipette. The agar tube with the embryo was expelled into a black grooved dish and covered with water and a coverslip.

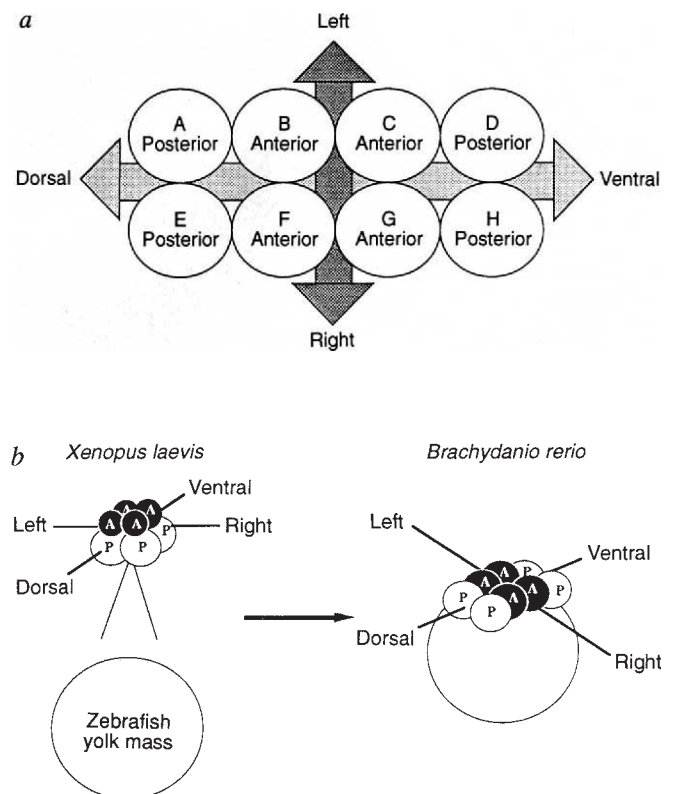


FIG. 2 a, Schematic depiction of the axes at the eight-cell stage. LR and DV axes are perpendicular and lie along the short and long sides of the 2 × 4 cell array. The AP axis is distributed such that the central blastomeres, B, C, F and G, project to the anterior structures of the adult. The outer blastomeres, A, D, E and H contribute to the posterior structures as described in the text. b, The axes of the *Xenopus laevis* fate map can be conceptually transformed into a zebrafish fate map. This can be accomplished by inserting a representation of the fish yolk between the posterior cells of *Xenopus* and spreading the dorsal posterior cells apart from the ventral posterior cells.

FIG. 3 Typical representations of the patterns of dye distribution from embryos injected at the eight-cell stage. The dorsal-ventral (DV) axis is represented by A; the left-right (LR) axis is represented by B; and the anterior-posterior (AP) axis by C. Aa, Ba and Ca show which blastomeres were injected to produce the patterns in each column. The DV pattern is represented in two separate animals in Ab, c and Ad, e, respectively. The animal in photographs Ab, c possesses a red ventral pattern and a green dorsal pattern. The animal in Ad, e, having been injected with colours in the opposite blastomeres, possesses the opposite pattern: a green ventral and a red dorsal pattern. The LR pattern is depicted by one animal in photographs Bb-e. The AP pattern is depicted by two animals in Cb, c and Cd, e. The animal depicted in panel Cb, c was injected in neighbouring blastomeres on the ventral and right sides of the eight-cell stage embryo, with the green dye injected in the outer, posterior blastomere. Hence the green pattern in Cc possesses much less label in the head. The animal depicted in Cd and Ce was injected in blastomeres in the dorsal and left sides of the embryo with the red dye injected in the outer, posterior blastomere. The outer blastomeres on the dorsal side of the eight-cell stage embryo contribute to the head structures up to but not including the diencephalon. Green epifluorescence was detected with an exciter filter of 450–490 nm and a barrier filter of 520–560 nm. For red the exciter filter was 546 nm and the barrier filter was 590 nm. Images in Aa, Ba, C, Ad, e and Bd, e were double exposures. Total magnifications are as follows: Ab, c, Bb–d, Cb, Cc: $\times 18$; Aa, Ba, Ca: $\times 28$; Ad, e, Be, Cd, e: $\times 47$.

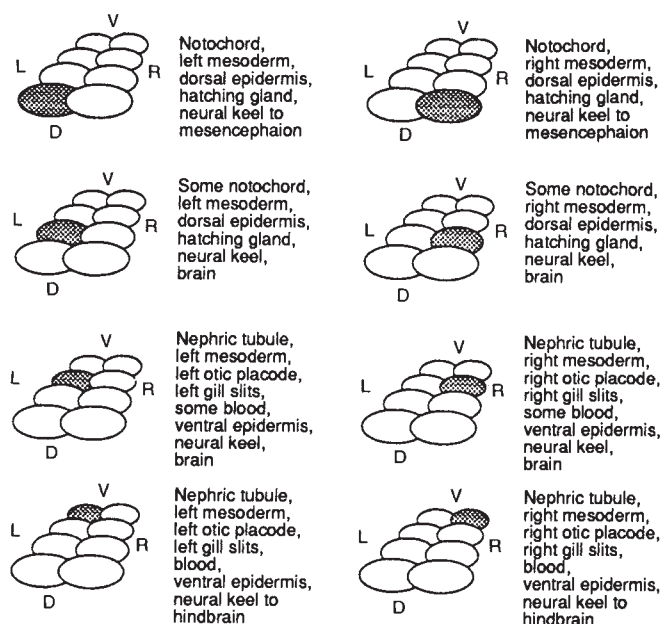
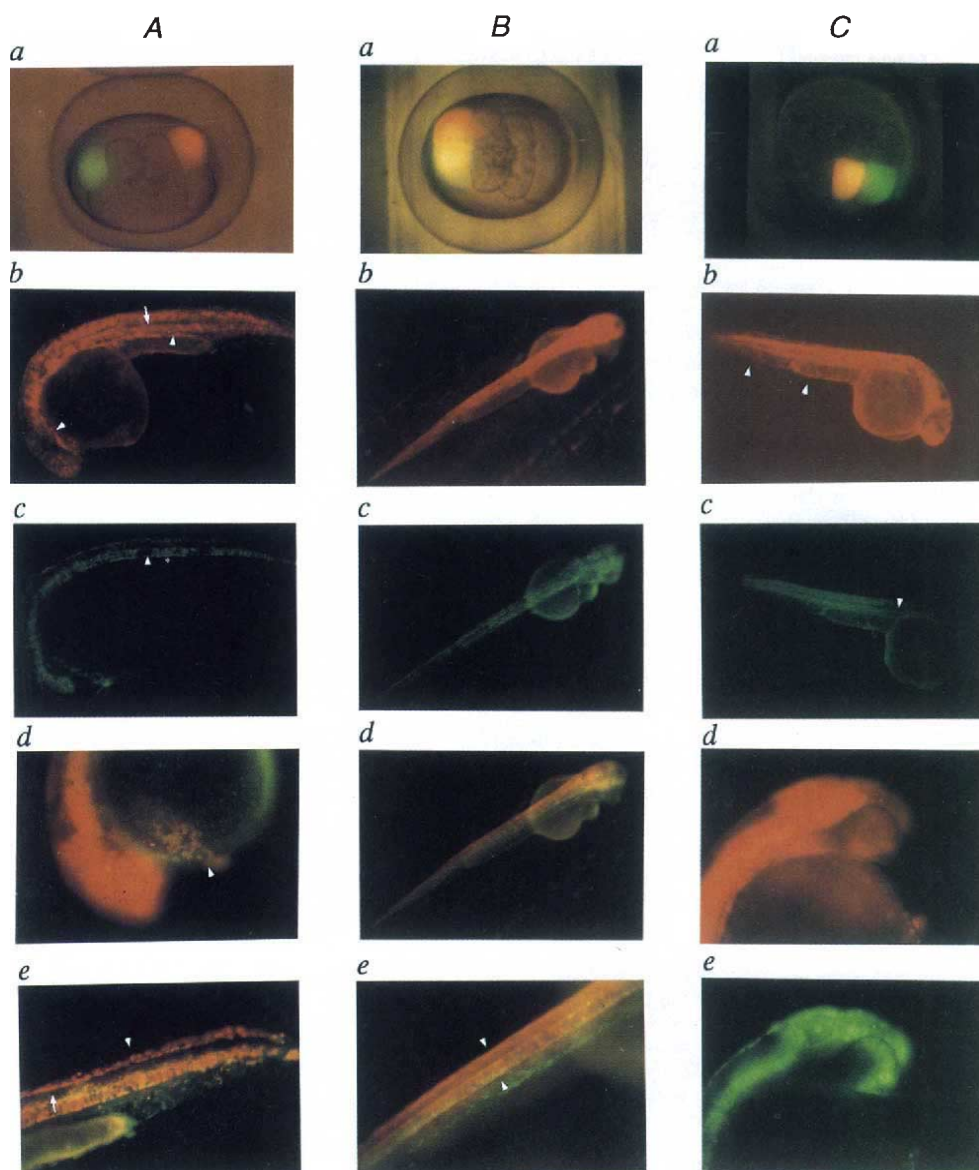


FIG. 4 The contribution of each eight-cell blastomere to the various tissues. Note the retina is labelled occasionally by each blastomere; the olfactory placodes are labelled occasionally with the dorsal epidermis and occasionally with the ventral epidermis.

of the spinal cord occurs to a small degree in both the dorsal and ventral patterns.

Figure 3B shows the left-right (LR) segregation of lineages. The differences between labelling on the left or right sides is seen in the mesoderm contributing to the somite muscle and also in the primordial ear (the otocyst) and primordial gill slits. All of the red fluorescent muscle is restricted to the left side of the animal, and all of the green fluorescent muscle is on the right side (arrowheads in *Be*). This appears in a double exposure in *Bd* and *Be*. Both blastomeres give rise to separate cells in the notochord, causing it to appear yellow in *Bd* and *Be*. The dorsal blastomeres, as mentioned above, also contribute significantly to the head (up to the diencephalon) although this is not readily visible in the figure.

Figure 3C shows the differences in lineages specified along the anterior-posterior (AP) axis. The central cells in the 2×4 cell array contribute to the head, whereas the outer cells contribute to the body plan from the tail to the hindbrain (if on the ventral side of the array) and from the tail to but not including the telencephalon (if on the dorsal side). In *Cb*, the red fluorescence extends through the head, whereas the green fluorescence in *Cc* extends from the tail to the hindbrain (arrowhead in *Cc*). The ventral patterns correspond to that described in *A*, above, and the right side of each embryo is labelled corresponding to the proposed LR axis. The embryo depicted in *Cd* and *Ce* (reversed label: red dye into the posterior blastomere and green dye anterior) shows the more subtle difference between the anterior and posterior blastomeres when they are located on the dorsal side of the DV axis. The posterior label extends from the tail through the mesencephalon. The anterior label, in green, extends completely through the brain.

About 90% (82 out of 92) of the blastomeres displayed lineage patterns which agreed with the postulated axes (Fig. 2a). Those which did not show three correct axes differed either because the orientation of the LR axis did not correlate with the orientation of the DV axis (three embryos) or because dorsal lineages arose from ventral blastomeres (two embryos). That the frequency of success is not closer to 100% may reflect a technical difficulty in correctly injecting the early blastomeres without damaging them, or it may reflect a natural variability in the lineages.

For most tissues, including especially the notochord, floorplate, dorsal and ventral epidermis, hatching gland, mesoderm, blood, and ectoderm, we see reliable lineage restrictions. But a few tissues, notably the retina, the lens and the olfactory placode, involve the contributions of a variety of progenitors. Nonetheless, the general form of the patterns is highly reproducible. Figure 4 summarizes the contributions of each eight-cell blastomere to the various tissues.

The existence of the *Xenopus* fate map provokes comparison with this new data. Figure 2b shows that the pattern of the *Xenopus* fate map can be conceptually transformed into that of the zebrafish by a simple manoeuvre: inserting a large zebrafish yolk mass directly under the *Xenopus* vegetal cells.

This demonstration of extremely similar fate maps for the fish and the amphibians strongly suggests that this pattern of well defined fate maps extending from the first cleavages will turn out to be characteristic of vertebrate development. □

Tissue-plasminogen activator is induced as an immediate-early gene during seizure, kindling and long-term potentiation

Zhuo Qian, Mary E. Gilbert*, Michael A. Colicos, Eric R. Kandel & Dietmar Kuhl

Center for Neurobiology and Behavior, and Howard Hughes Medical Institute, College of Physicians and Surgeons of Columbia University, 722 West 168th Street, New York, New York 10032, USA
* ManTech Environmental Technology Inc., PO Box 12313, Research Triangle Park, North Carolina 27709, USA

THE requirement of protein and messenger RNA synthesis for long-term memory^{1,2} suggests that neural activity induced by learning initiates a cascade of gene expression³. Here we use differential screening to identify five immediate-early genes induced by neuronal activity. One of these is tissue-plasminogen activator (tPA), an extracellular serine protease, which is induced with different spatial patterns in the brain by three activity-dependent events: (1) convulsive seizure increases expression of tPA in the whole brain; (2) stimulation of the perforant path produces an epileptiform after-discharge that ultimately leads to kindling increases the levels of tPA throughout the hippocampus bilaterally; and (3) brief high-frequency stimulation of the perforant path that produces long-term potentiation (LTP) causes an NMDA (*N*-methyl-D-aspartate) receptor-mediated increase in the levels of tPA mRNA which is restricted to the granule cells of the ipsilateral dentate gyrus. As release of tPA is correlated with morphological differentiation⁴⁻⁶, the increased expression of tPA may play a role in the structural changes that accompany activity-dependent plasticity⁷⁻¹⁰.

To identify new immediate-early genes whose expression is dependent on neuronal activity, we differentially screened ~30,000 primary clones from our hippocampal complementary DNA library (see Fig. 1 legend) using two cDNA probes derived from mRNA isolated from the cortex of untreated animals (– probe) or from the cortex of animals that had undergone metrazol-induced seizure in the presence of protein synthesis inhibitor (+ probe). Clones that hybridized preferentially to the (+) probe were further investigated by slot blot and northern blot hybridization. We identified, in addition to *fos* and *zif268*, five independent clones initially named BAD1–5 (for brain-activity-dependent). Partial sequence analysis of these clones and search of the Genbank database revealed that BAD1, BAD2, and BAD3 are novel genes. BAD4 encodes a putatively secreted protein (PC3 or TIS21) of unknown function that is activated by nerve growth factor (NGF), serum, and seizure activity^{11,12}. We found that BAD5 is the tissue-type plasminogen activator (tPA), an extracellular serine protease that converts plasminogen into plasmin, a broad spectrum protease.

Although they differ in their patterns of induction by seizure activity (Fig. 1A), all five genes have characteristics shared by immediate-early genes. The basal levels of the mRNAs are low or undetectable, but they can be rapidly induced. Moreover, the mRNA transcripts continue to accumulate in the presence of the protein synthesis inhibitor cycloheximide (Fig. 1A), indicating that the induction of each of these genes by seizure activity is independent of new protein synthesis. Finally, as is the case with *c-fos* and several other immediate-early gene transcripts, we have observed that the mRNA of tPA contains several repeats of AUUUA in its 3' untranslated region; this sequence may contribute to the instability of tPA transcripts¹³.

The immediate-early genes so far found to be inducible by electrical activity in nerve cells were all initially discovered in non-neuronal cells¹⁴⁻¹⁸. We therefore examined the expression

Received 3 September; accepted 19 November 1992.

- Jacobson, M. & Horose, G. *J. Neurosci.* **1**, 271–284 (1991).
- Jacobson, M. & Moody, S. A. *J. Neurosci.* **4**, 1361–1369 (1984).
- Dale, L. & Slack, J. M. *Development* **99**, 527–551 (1987).
- Moody, S. A. *Dev. Biol.* **122**, 300–319 (1987).
- Kimmel, C. B. & Law, R. D. *Dev. Biol.* **108**, 86–93 (1985).
- Kimmel, C. B. & Law, R. D. *Dev. Biol.* **108**, 94–101 (1985).
- Kimmel, C. B. & Warga, R. M. *Science* **231**, 365–368 (1986).
- Kimmel, C. B. & Warga, R. M. *Dev. Biol.* **124**, 269–280 (1987).
- Kimmel, C. B., Warga, R. M. & Schilling, T. F. *Development* **108**, 581–594 (1990).
- Warga, R. M. & Kimmel, C. B. *Development* **108**, 569–580 (1990).
- Kimmel, C. B. & Law, R. D. *Dev. Biol.* **108**, 78–85 (1985).
- Debelder, A. N. & Dranath, K. *Carb. Res.* **30**, 375 (1973).

of the BAD1-5 genes in fibroblasts. Serum stimulation increased transcript levels of all five genes; however, the kinetics of induction in some cases differed from those in the brain (Fig. 1B). Most striking is that expression of tPA in the brain peaks at 1 h after seizure, whereas in fibroblasts it is still increasing even 4 h after serum stimulation. The different patterns in the expression of these genes might reflect the distinct phenotypic responses of different cells to different stimuli.

As tPA is a distinctive immediate-early gene encoding an extracellular serine protease, it seemed likely to be eventually accessible for study using specific inhibitors and antibodies. We therefore focused on tPA and studied its response to different forms of stimulation. First, we examined the distribution of tPA expression by metrazol-induced seizure activity. Using *in situ* hybridization, we found that 1 h after the onset of seizure, the mRNA for tPA is increased about sixfold throughout the brain. The levels of tPA mRNA remained elevated for at least 4 h (Fig. 2).

We next examined kindling, a form of experimental epilepsy used as a model of neuronal plasticity, because, like LTP, it leads to a prolonged enhancement of synaptic strength in the stimulated pathway and shares with LTP a number of physiological and biochemical characteristics¹⁹⁻²². In kindling, a stimulus that initially evokes only a brief after-discharge which is subthreshold for seizure becomes capable, as a result of repetition, of consistently initiating a generalized seizure. Once established, this hyperexcitable state endures for the life of the animal. To produce an after-discharge that leads to kindling, we stimulated the perforant path through indwelling electrodes (60 Hz for 1 s using 1-ms pulses of 500 μ A; $n = 3$). This stimulation, which induced an after-discharge in the granule cells but no behavioural seizure, caused strong bilateral induction of tPA mRNA at 1 h throughout the hippocampus (Fig. 3). When we stimulated below threshold for evoking an after-discharge (100 μ A), we did not observe induction of mRNA levels for tPA ($n = 2$).

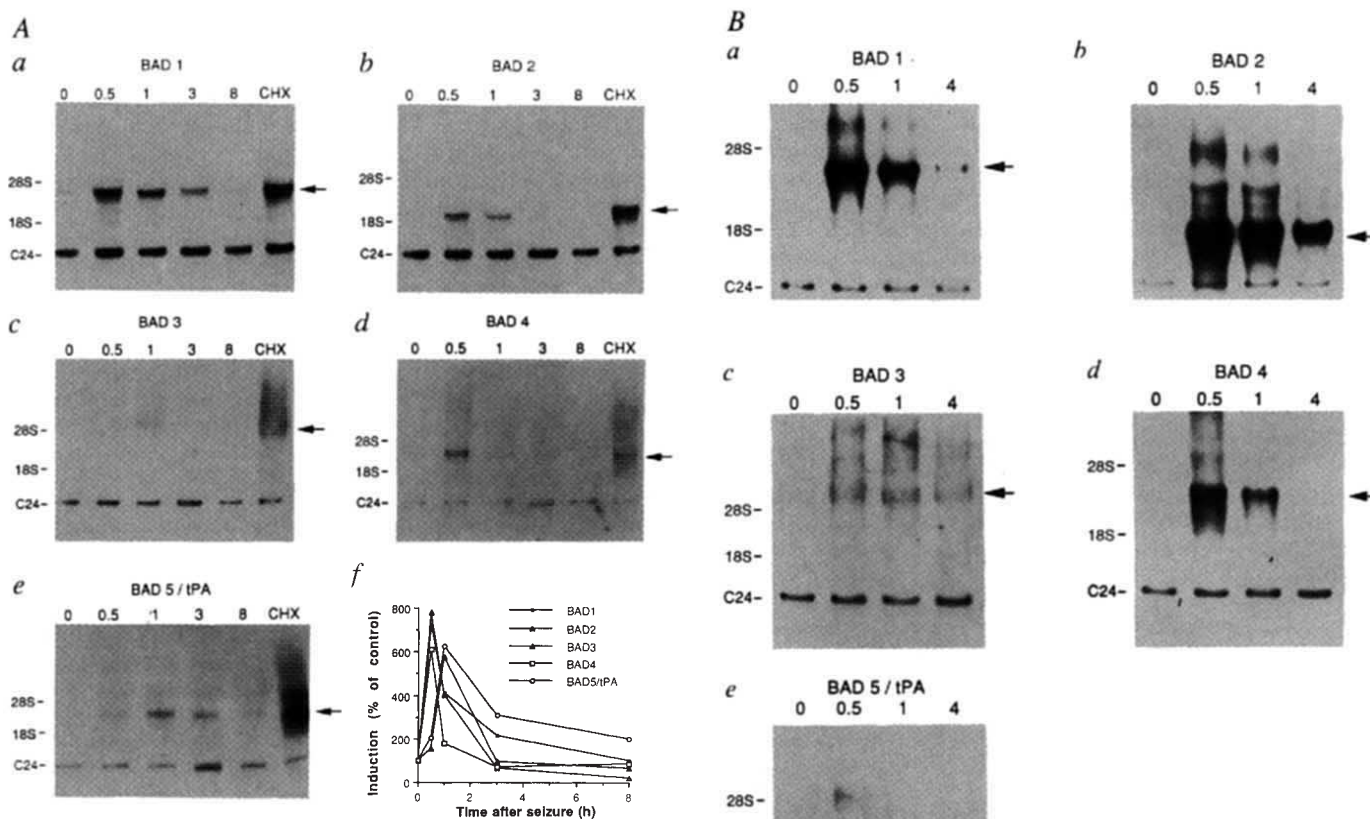


FIG. 1 A, Induction and decay of the levels of five brain activity-dependent mRNAs (BAD 1-5) after metrazol-induced seizure activity. a-e, Autoradiographs of RNA blots. Total cellular RNA from the cortex of control animals and animals that had undergone metrazol-induced seizure either in the presence or absence of the protein synthesis inhibitor cycloheximide was isolated at different times and used in RNA blot analysis (5 μ g per lane). The blots were hybridized to probes specific for BAD 1-5. Hybridization to a probe specific for C24, previously isolated from the hippocampal cDNA library on the basis of its lack of response to seizure, was used as a loading control. Numbers above the lanes indicate time in hours after the onset of seizure at which RNA was isolated. Lanes labelled 'CHX' contain RNA isolated 3 h after the onset of seizure in the presence of cycloheximide. The positions of BAD 1-5, C24, 28S and 18S rRNAs are indicated. f, to quantify the RNA levels after seizure, the autoradiographs were scanned and results are plotted. B, Kinetics of induction of BAD 1-5 in fibroblasts after serum stimulation. Autoradiography of RNA blots. Total cellular RNA (3 μ g) from resting or serum stimulated Rat-6 fibroblasts were used in each lane of the RNA blot analysis. Probes and abbreviations are as in A. Serum induction was as described²⁹. Similar results were obtained with mouse NIH 3T3 cells (data not shown).

METHODS. For the construction of the seizure-induced cDNA library, rats were injected intraperitoneally with cycloheximide (120 μ g per kg body weight), 30 min later the rats were injected with metrazol (50 μ g per kg body weight) to produce a seizure. One hour after the onset of seizure, animals were decapitated and poly(A)⁺ RNA was isolated (Pharmacia) from the hippocampus and used to generate an oligo(dT)-primed cDNA library in the plasmid vector pcDNAII (Invitrogen). Thirty thousand clones from this library were differentially screened²⁸, using ³²P-labelled cDNA probes derived from poly(A)⁺ RNA isolated from the cortex of either animals that had undergone metrazol-induced seizure in the presence of cycloheximide or from untreated animals. RNA blot analysis with random-primed cDNA inserts (Pharmacia) was as described²⁷. For quantification autoradiographs were scanned with a Molecular Dynamics 100-A laser densitometer and analysed with Molecular Dynamics Software. The optical densities of each specific signal for BAD 1-5 in a to e were normalized with the corresponding value obtained for the C24 reference band.

Finally, we studied LTP. To evoke LTP in the granule cells, we delivered to the perforant path, in freely moving rats, ten high-frequency trains (400 Hz for 50 ms using 0.1-ms pulses of 1,000 μ A). This high-frequency stimulation produced LTP that lasted for at least 24 h, as evidenced in a significant increase in the field potential (Fig. 4a). The high-frequency train also caused a rapid and marked increase at 1 h in the level of tPA mRNA. But, in contrast to stimulation leading to kindling, stimuli that evoke LTP produced an induction of tPA, which is restricted to the granule cell neurons of the ipsilateral dentate gyrus (Fig.

4c; LTP, 1 h; $n = 4$). In response to stimuli that produce LTP, we also observed a pattern of increase in the level of the transcription factor *zif268* (data not shown) that was similarly restricted to the granule cells. This molecular marker is thought to be specific to LTP^{16,18}. To ensure further that the changes in tPA mRNA are only related to LTP and not to adventitious after-discharge, we also monitored throughout the experiment the electrical activity in the electroencephalograms (EEGs). We detected no after-discharge or seizure activity during or after the initiation of LTP.

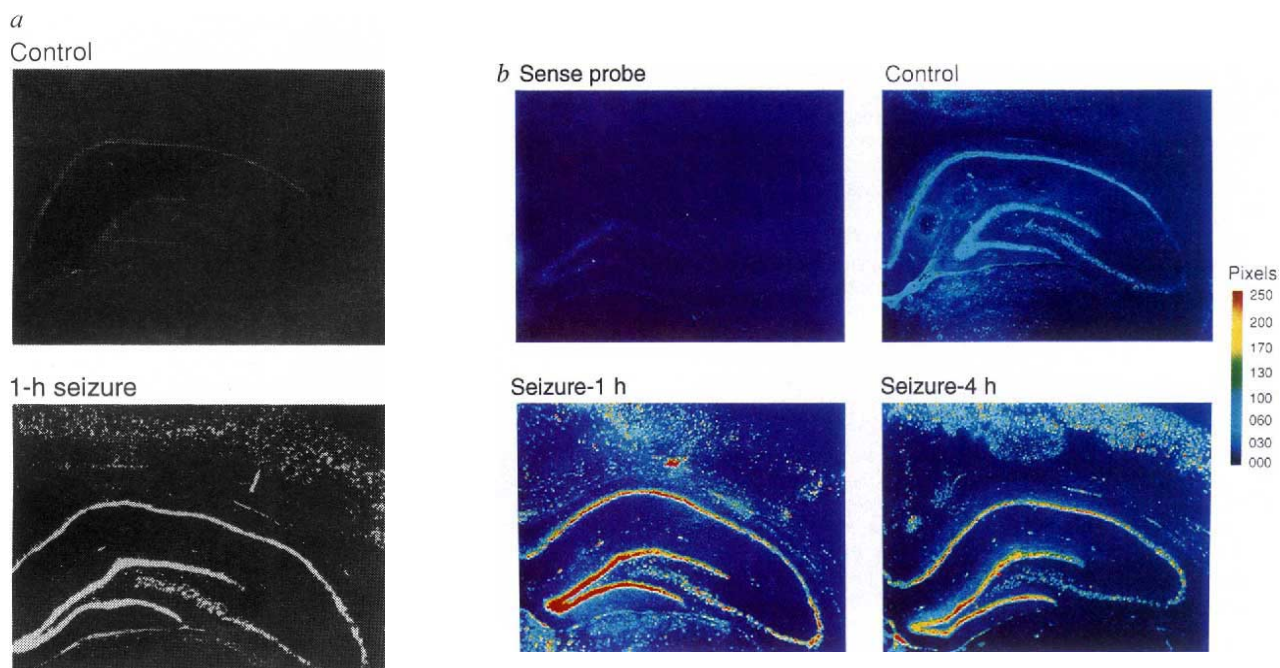


FIG. 2 Localization by *in situ* hybridization of tPA mRNA expression after seizure. *a*, Representative *in situ* hybridization data before pseudocolour image conversion, before (control) and 1 h after seizure. *b*, Pseudocolour representation of tPA mRNA expression, before (control) and at 1 and 4 h after seizure.

METHODS. Tissue sections through the rostral hippocampus of control rats and rats seized for various hours were prepared⁸. Antisense- and sense-specific tPA RNA probes were labelled with digoxigenin-UTP (RNA Labeling Kit, Boehringer Mannheim). *In situ* hybridizations were done at 42 °C overnight. Slides were washed in 2 $\times$ SSC, 1 $\times$ SSC, each for 1 h at room temperature, and then in 0.3 $\times$ SSC at 37 °C for 30 min and finally at room

temperature in 0.3 $\times$ SSC for an additional 30 min. Slides were then air-dried and incubated with an anti-digoxigenin antibody conjugated to fluorescein (20 μ g ml⁻¹) for 1 h at room temperature. Slides were briefly washed three times in 1 $\times$ PBS and examined by computer-linked fluorescent microscopy (Zeiss IM35). Emitted fluorescence was recorded through a computer-linked camera. To facilitate the quantification, black-white *in situ* pictures were converted to pseudocolour. Quantification of fold induction is achieved by comparison of pixel values of an area of interest before and after seizure. (Similar results were obtained in at least two additional experiments on different animals.)

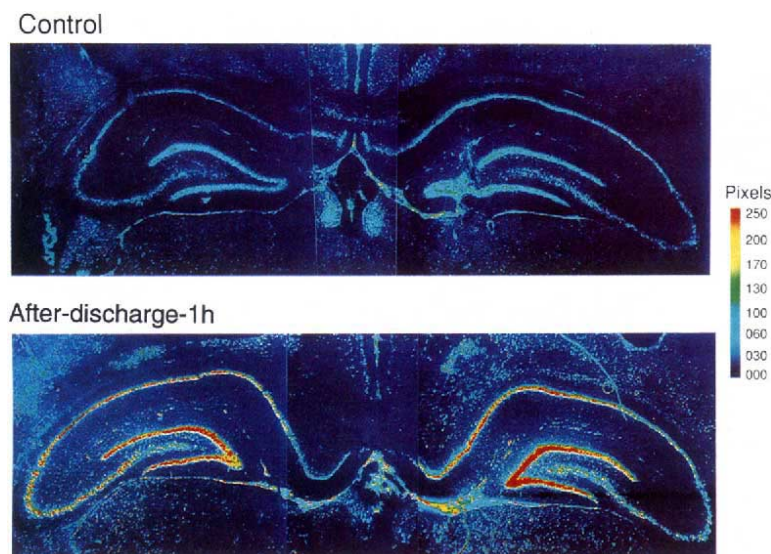


FIG. 3 Elevation of tPA mRNA in hippocampus following induction of after-discharge. Five rats were prepared with chronic indwelling electrodes in the angular bundle of the perforant path and hilus of the dentate gyrus²⁴. Following a 2-3 week recovery period, three animals received a suprathreshold train (60 Hz for 1 s using 1-ms pulses at 500 μ A) that induced a mean of 34 s of after-discharge in the dentate gyrus. No behavioural seizures were observed. mRNA for tPA was elevated throughout the hippocampus bilaterally in these animals. The other two animals were administered a lower stimulus subthreshold for after-discharge (100 μ A) and no alteration in tPA mRNA was observed. *In situ* hybridization was carried out as described in the legend to Fig. 2.

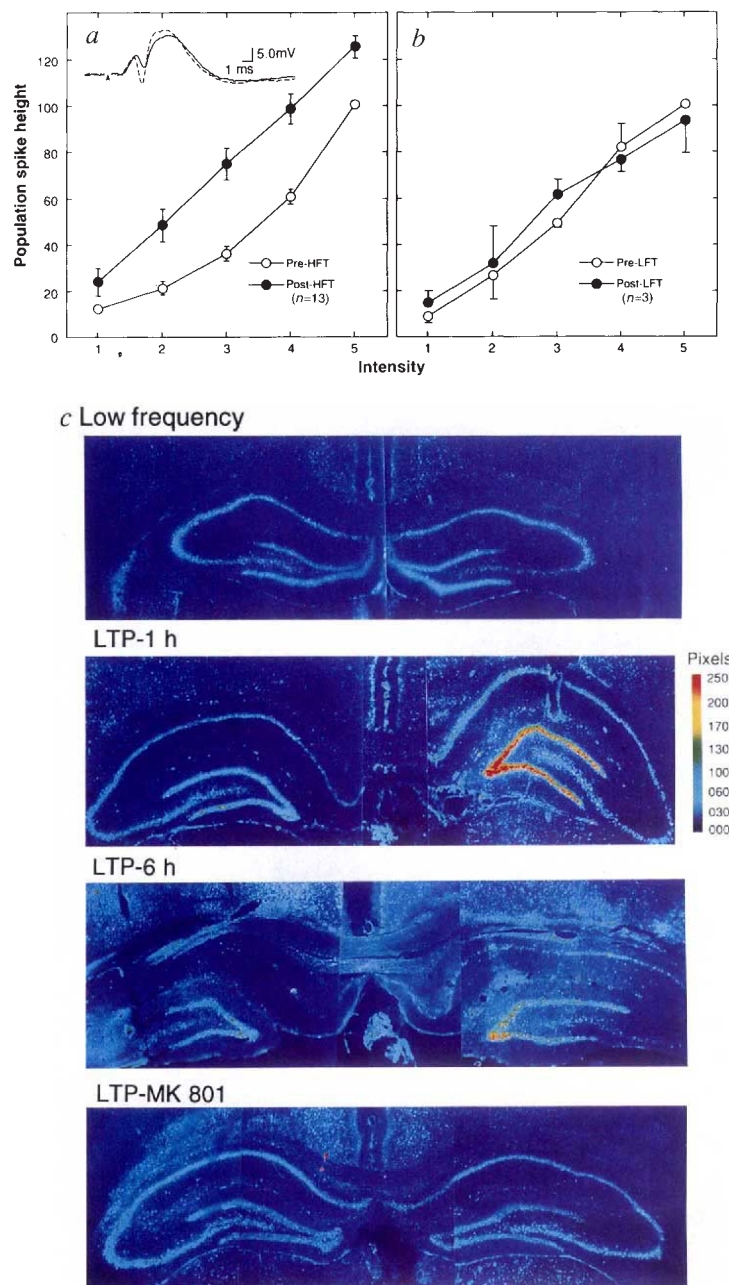


FIG. 4 Induction of LTP and tPA mRNA in the perforant path-dentate gyrus of freely moving rats. In *a*, high-frequency stimulation evokes LTP in freely moving rats, whereas in *b*, low-frequency stimulation does not evoke LTP in freely moving rats. HFT, high-frequency train. *c*, tPA mRNA levels in stimulated animals. LTP-low frequency and LTP-MK801 means 1 h after application of low-frequency stimulation and 1 h after high-frequency stimulation with pre-injection of MK801 to block induction of LTP, respectively.

METHODS. Sixteen rats, prepared as described in Fig. 3 legend, were given a series of biphasic square-wave stimulus pulses delivered to the perforant path and the response evoked in the dentate gyrus was recorded. An input/output function, comprising ten pulses (1 pulse per 10 s) delivered at each of 5 stimulus intensities (20–1,200 μA), was collected before (open symbols) and 45–60 min following (filled symbols) the administration of 5 pairs (2 s apart) of high-frequency trains (50 ms, 400 Hz, 0.1 ms pulse width, 1,000 μA at 2-min intervals). Animals were killed 1 h ($n=4$), 6 h ($n=4$), or 24 h ($n=3$) after train delivery and brain tissue was prepared for *in situ* hybridization as described in Fig. 2 legend. The inset in *a* shows the response evoked by a 150-μA pulse in one animal before (solid line) and 50 min after (broken line) HFT. The elevation of mRNA for tPA in the granule cells of the ipsilateral dentate gyrus of this animal is shown in *c*, LTP-1 h. In addition, three LTP-control animals receiving the same number of stimulus pulses delivered at a low frequency (LFT, 0.05 Hz; *b*) and two animals administered MK801 (2 mg per kg; data not shown) 45 min before delivery of HFT failed to demonstrate an increase in population spike height 45–60 min post-train. These five animals were killed 1 h post-train delivery, and prepared for *in situ* hybridization.

Following its initial induction with LTP, the levels of tPA mRNA decreased by 6 h ($n=4$; Fig. 4c; LTP, 6 h) and returned to basal levels by 24 h ($n=3$; data not shown). Both LTP and the increase in the level of tPA mRNA were blocked by pre-injection of the non-competitive NMDA receptor antagonist, MK801 ($n=2$), suggesting that the induction of LTP and tPA transcripts is correlated with and requires the activation of NMDA receptors. Similarly, the induction of LTP and of the tPA transcript seemed to have a comparable threshold. Delivery of the same number of stimulus pulses at a low frequency (0.05 Hz; $n=3$), which did not evoke LTP (Fig. 4b), also did not alter the mRNA levels of tPA (Fig. 4c; LTP low-frequency).

These data indicate that three forms of neural activity (seizure, evoked after-discharge which leads to kindling, and LTP) lead to the induction of tPA mRNA with three distinct spatial patterns of expression. The increase in mRNA levels for tPA is likely to be reflected by a concomitant elevation of tPA protein released into the extracellular space, because studies in cell culture show that increases in tPA mRNA always precede gradual increases

in cellular tPA and its subsequent secretion into the medium²³. Interestingly, induction of LTP is associated with a delayed increase in an unidentified extracellular protease of a molecular weight similar to that of tPA²⁴.

What role might tPA have in these three forms of neuronal plasticity? Stimuli that evoke seizure, kindling, or LTP are all capable of inducing structural changes^{7,10}. How neuronal activity leads to the morphological changes is not known. As tPA is released from the growth cones⁴ and this release is correlated with morphological differentiation^{5,6}, induction and release of tPA could contribute to the initiation of activity-related structural changes in one of three ways: (1) by altering adhesive contacts between neurons^{6,7}; (2) by changing the spatial and temporal interaction between proteases and protease inhibitors²⁵; or perhaps (3) by activating a potential receptor²⁶ in a manner similar to that activated by thrombin²⁷, a serine protease. The finding of a neuronal, activity-dependent, immediate-early effector gene of known function, whose protein acts extracellularly, now allows use of specific inhibitors and anti-

bodies to explore how this protein contributes to the maintenance of LTP and of kindling and to the structural changes that accompany them. □

Received 11 September; accepted 8 December 1992.

- Davis, H. P. & Squire, L. R. *Psychol. Bull.* **96**, 518–519 (1984).
- Montarolo, P. G., Goelet, P., Castellucci, V. F., Morgan, J. I. & Kandel, E. R. *Science* **234**, 1249–1254 (1986).
- Goelet, P., Castellucci, V. F., Schacher, S. & Kandel, E. R. *Nature* **322**, 419–422 (1986).
- Krystosek, A. & Seeds, N. W. *Science* **213**, 1532–1533 (1981).
- Neuman, T., Stephens, R. W., Salonen, E.-M., Timmusk, T. & Vaheri, A. J. *Neurosci. Res.* **23**, 274–281 (1989).
- Pittman, R. N., Ivins, J. K. & Buettner, H. M. *J. Neurosci.* **9**, 4269–4286 (1989).
- Ben-Ari, T. & Represa, A. *Trends Neurosci.* **13**, 312–318 (1990).
- Cavazos, J. E., Galarai, G. & Sutula, T. P. *J. Neurosci.* **11**, 2795–2803 (1991).
- Wallace, C. S., Hawrylyk, N. & Greenough, W. T. in *Long-Term Potentiation* (eds Baudry, M. & Davis, J. L.) 189–222 (MIT Press, Cambridge, MA, 1991).
- Geinisman, Y. et al. *Brain Res.* **569**, 342–347 (1992).
- Bradbury, A., Possenti, R., Shooter, E. & Tirone, F. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3353–3357 (1991).
- Fletcher, B. S. et al. *J. Biol. Chem.* **266**, 14511–14518 (1991).
- Shaw, G. & Kramer, R. *Cell* **24**, 151–260 (1986).
- Greenberg, M. E., Greene, L. A. & Ziff, E. B. *J. Biol. Chem.* **260**, 14101–14110 (1985).
- Morgan, J. I., Cohen, D. R., Hempstead, J. L. & Curran, T. *Science* **237**, 192–197 (1987).
- Cole, A. J., Saffen, D. W., Barban, J. M. & Worley, P. F. *Nature* **340**, 474–476 (1989).
- Dragunow, M. & Robertson, H. A. *Nature* **229**, 441–442 (1987).
- Wisden, W. et al. *Neuron* **4**, 603–614 (1990).
- Racine, R. et al. *J. Electroencephalogr. clin. Neurophysiol.* **32**, 269–279 (1972).
- Racine, R. J., Milgram, N. W. & Hafner, S. *Brain Res.* **260**, 217–231 (1983).
- McNamara, J. O., Byrne, M. C., Dashieff, R. M. & Fitz, J. G. *Prog. Neurobiol.* **15**, 139–159 (1980).
- Gilbert, M. E. & Mack, C. M. *Brain Res.* **519**, 89–96 (1990).
- Gerard, R. & Meidell, R. S. A. *Rev. Physiol.* **51**, 245–262 (1989).
- Faselli, M. S., Errington, M. L., Dolphin, A. C. & Bliss, T. V. P. *Brain Res.* **521**, 247–253 (1990).
- Monard, D. *Trends Neurosci.* **11**, 541–544 (1988).
- Hajjar, K. A. *J. Biol. Chem.* **266**, 21962–21970 (1991).
- Carney, D. H. *Proteases in Biological Control and Biotechnology* (ed. Cunningham, D. D. & Long, G. L.) 277–282 (Liss, New York, 1987).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Almendral, J. et al. *Molec. cell. Biol.* **8**, 2140–2148 (1988).

ACKNOWLEDGEMENTS. We thank J.-F. Brunet, P. Dash and H. Coles for their help in a pilot study. S. Nichols and T. Livelli for rat-6 and mouse NIH 3T3 fibroblasts, and M. Osman for assistance with animal experiments. E.R.K. is a Senior Investigator and D.K. and Z.O. are Research Associates of the Howard Hughes Medical Institute. This research was supported by the Howard Hughes Medical Institute, by grants from the NIA and the NIMH, and by the US Environmental Protection Agency to ManTech Environmental Technology, Inc.

Assembly of GABA_A receptor subunits determines sorting and localization in polarized cells

Jose L. Perez-Velazquez & Kimon J. Angelides*

Departments of Cell Biology, Neuroscience and Biochemistry,
Baylor College of Medicine, Houston, Texas 77030, USA

THE GABA_A receptor, the principal inhibitory receptor in the CNS, is distributed on cell bodies, dendrites, and in some cells at axon hillocks and presynaptic terminals^{1–6}. The dendritic distribution is crucial for shunting of excitatory synaptic inputs^{7,8}. Molecular cloning has revealed that the GABA_A receptor can be formed by a diverse set of subunits and by separately encoded subunit isoforms^{9,10}, the expression of each of which differs in distinct areas of the central nervous system^{11–13} and during development^{14,15}. Why different genes exist to encode these isoforms is not clear, but may be linked to functional differences^{16–19}. Here we show that assembly of specific isoforms also codes for sorting and localization of the receptor complex. Confocal microscopy and immunoblot analysis of epithelial cells transfected with the complementary DNAs encoding the $\alpha 1$ and $\beta 1$ GABA_A receptor subunits and probed with subunit isoform-specific antibodies show that the $\alpha 1$ subunit is targeted to the basolateral surface, and that the $\beta 1$ subunit is sorted to the apical membrane. In cells where $\alpha 1$ and $\beta 1$ isoforms are co-expressed,

assembly of the $\beta 1$ with the $\alpha 1$ subunit isoform re-routes the $\alpha 1$ subunit to the apical surface. The ability to assemble complexes of different isoform composition and to target these to specific regions of the cell surface would enable neurons to modulate GABA_A receptor distribution and possibly alter the composition of its synapses in response to transcriptional levels of specific subunit isoforms.

Studies on the sorting of viral proteins^{20,21} and isoforms of neural cell adhesion molecules²² in epithelial cells and hippocampal pyramidal neurons suggest that there are similarities in the mechanisms by which neurons and epithelial cells sort and construct their plasma membranes²³. We have taken advantage of these similarities to examine the signals that mediate sorting of GABA_A receptor subunits. MDCK cells transfected with cDNAs encoding either the $\alpha 1$ or $\beta 1$ subunit isoforms, placed under the control of the constitutive SV40 promoter, are able to transcribe both $\alpha 1$ and $\beta 1$ subunit messenger RNAs (Fig. 1).

The localization of $\alpha 1$ and $\beta 1$ subunits was examined by immunocytochemistry using $\alpha 1$ - and $\beta 1$ -specific antibodies. Confocal micrographs of $\alpha 1$ -transfected cells showed that the immunofluorescence was localized to the basolateral membrane (Fig. 2a). Optical sections taken at the level of the apical and

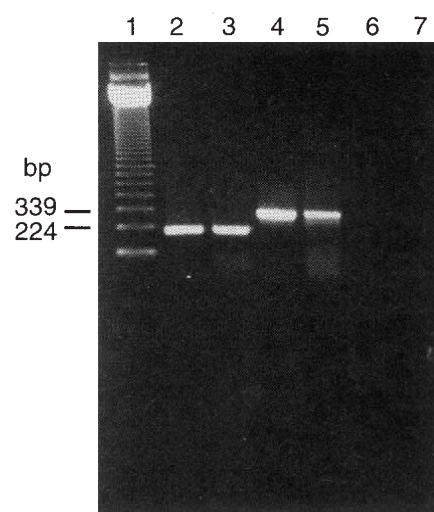


FIG. 1 Transcription of $\alpha 1$ and $\beta 1$ GABA_A receptor mRNAs in transfected MDCK cells. Two conserved oligonucleotides were chosen spanning a 224-bp region of the $\alpha 1$ bovine sequence and 339 bp of the $\beta 1$ bovine sequence. These primers were used in the polymerase chain reaction (PCR) using total RNA isolated from transfected and control cultures. Lane 1, 123-bp DNA ladder; lane 2, PCR fragment from plasmid encoding the $\alpha 1$ subunit in pSP64 (gift from P. Seeburg); lane 3, PCR product from $\alpha 1$ -transfected cultures; lane 4, PCR product from $\beta 1$ cDNA in pSP64; lane 5, PCR fragment of $\beta 1$ -transfected MDCK cultures; lanes 6 and 7, PCR products from mock-transfected control cultures.

METHODS. MDCK cells were obtained from E. Rodriguez-Boulan. Cells were plated in MEM (minimum essential Eagle's medium), with 10% fetal bovine serum (FBS). Plasmid DNA for transfection was prepared by the alkaline lysis method followed by polyethylene glycol precipitation or by extraction with the SpinBind kit (FMC BioProducts). Epithelial cells were plated at a density of 2×10^5 cells per 35-mm dish (coated with polylysine) 48 h before transfection. Each 35-mm dish of confluent MDCK cells was transfected with 5–10 μ g of the pSVK3- $\alpha 1$ or pSVK3- $\beta 1$ recombinant plasmid and 20 μ l (20 μ g) lipofectin (BRL) for 4–5 h at 37 °C in serum-free medium. Control cultures that lacked the GABA_A subunit cDNA (pSVK3 alone) were prepared by an identical transfection protocol. Total RNA was isolated from 35-mm dishes of transfected and controls cells 24 h post-transfection. First-strand cDNA was prepared from 1 μ g total RNA using random primers and Moloney murine leukaemia virus reverse transcriptase (BRL). The product from the first-strand synthesis was used in the polymerase chain reaction. PCR experiments were done with the following primers: $\alpha 1$: 5' end, ATGGG-ATCCGAAAAGTGTGGTCCAGAAAA; 3' end, TTGCTGCAGCTGTAAAGGT-TTCTTCGGT; $\beta 1$: 5' end, AAAGGATCCGGCAACAAGACCAGAGCGCTAAT; 3' end, CTGTCTGCAGAAATTCAGTCAGTCAAGTCGGGG.

* To whom correspondence should be addressed.

basolateral surfaces clearly show $\alpha 1$ staining only at the basolateral membrane and not the apical membrane. In addition, double-labelling with antibodies to $\alpha 1$ and spectrin, a protein segregated to the basolateral membrane of MDCK cells²⁴, showed that $\alpha 1$ was localized only on the spectrin-positive basolateral membrane.

In contrast, when MDCK cell cultures were transfected with $\beta 1$ cDNA and probed with an anti- $\beta 1$ antibody (Fig. 2c), confocal micrographs showed that the $\beta 1$ subunit was expressed on the apical domain, as confirmed by double-labelling with antibodies to the apical marker gp135 (Fig. 2d)²⁵. No staining was observed in optical sections at the level of the basolateral membrane in $\beta 1$ -transfected cells.

As $\alpha 1$ and $\beta 1$ subunits associate to form a functional channel when their corresponding cDNAs are expressed in oocytes and kidney cells^{16–19}, we asked whether, when they are associated, one of the subunits specifies sorting of the complex. MDCK cells were co-transfected with $\alpha 1$ and $\beta 1$ cDNAs and double-labelled with $\alpha 1$ - and $\beta 1$ -specific antibodies. Confocal micrographs obtained at both apical and basolateral membranes show that in those cells where both subunits are expressed, $\alpha 1/\beta 1$ complexes are localized exclusively to the apical membrane (Fig. 3a shows $\alpha 1$, and b shows $\beta 1$). In the field shown in Fig. 3b, one cell shows anti- $\beta 1$ on its apical side, characteristic of the location of the $\beta 1$ subunit, but lacks anti- $\alpha 1$ staining. In those cells in which both subunits are expressed, the $\alpha 1$ subunit was re-routed to the apical side in association with the $\beta 1$ isoform. Figure 3c and d are optical sections at the level of the basolateral membrane in this same field which show that there is no immunofluorescence associated with the expression of either $\alpha 1$ or $\beta 1$ on the basolateral membrane. Therefore, we

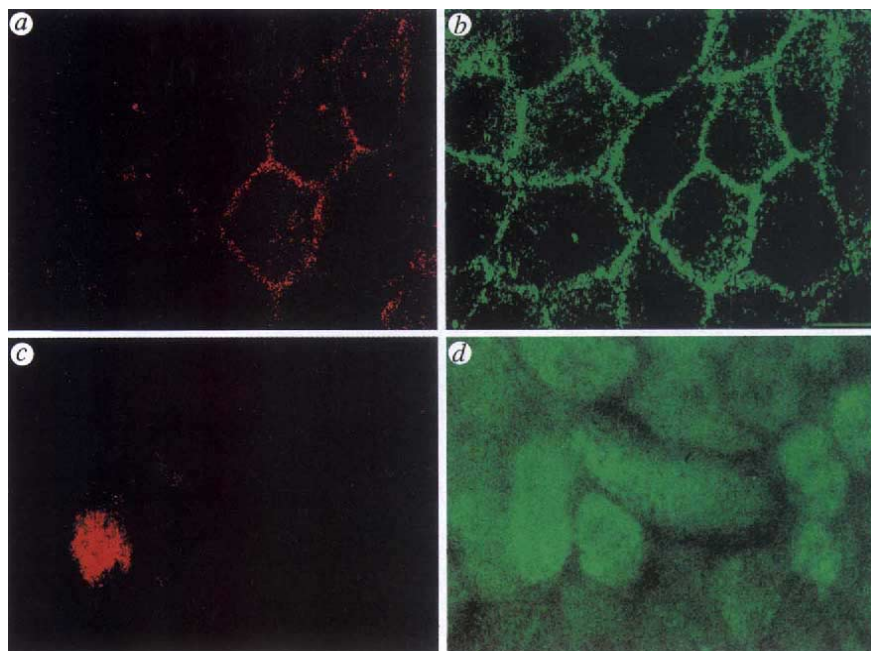
conclude that when the $\alpha 1$ and $\beta 1$ subunits assemble, the complex is targeted to the apical domain. In this way, $\beta 1$ re-routes the sorting of $\alpha 1$.

Finally, cells grown on permeable supports were biotinylated on either the apical or basolateral membranes, precipitated with streptavidin-Sepharose 6B, and then immunoblotted with both $\alpha 1$ - and $\beta 1$ -specific antibodies and with a polyclonal antibody against the purified GABA_A receptor²⁶. Consistent with immunofluorescence, cultures transfected with the $\alpha 1$ isoform express the $\alpha 1$ protein (M_r 50–53K) only on the basolateral side, with an additional band at 200K which possibly corresponded to an $\alpha 1$ tetramer (Fig. 4). In cultures transfected with $\beta 1$ cDNA, immunoblotting confirms the apical location of the $\beta 1$ subunit (Fig. 4). Immunoblotting also confirms the association of $\beta 1$ and $\alpha 1$ isoforms and the re-routing of the $\alpha 1$ subunit protein to the apical membrane by $\beta 1$ (Fig. 4). These results indicate that the $\alpha 1$ subunit proteins are located on the basolateral membrane and the $\beta 1$ subunit proteins are located on the apical membrane, whereas the $\alpha 1/\beta 1$ protein complexes are expressed on the apical surface.

These findings provide important insight into a potential consequence of subunit heterogeneity of the GABA_A receptor and have general implications for other multisubunit receptors such as those for glutamate and acetylcholine^{27,28}. In epithelial cells and neurons, the basolateral and dendritic pathways are considered to be default routes for protein sorting, whereas the apical/axonal route may require specific signals for protein targeting^{20,21}. Two different mechanisms have been proposed for sorting to apical membranes in epithelial cells and neurons^{20,21,23,29–32}: one in which the segregation of newly synthesized proteins takes place in the *trans*-Golgi through

FIG. 2 Immunofluorescence localization of GABA_A receptor $\alpha 1$ and $\beta 1$ subunits in transfected MDCK cells. a, Group of cells stained with anti- $\alpha 1$ antibody, optically sectioned at the level of the basolateral membrane. b, Same field co-stained with an antibody against the basolateral marker spectrin. There was no fluorescence detected at the apical membrane in this field. c, Cells expressing the $\beta 1$ subunit stained with an anti- $\beta 1$ -specific antibody. d, Same group of cells co-stained with a monoclonal antibody against the apical marker gp135, confirming the apical localization of the $\beta 1$ subunit. No fluorescence was detected at the basolateral membrane in this field. More than 500 cells were examined, from which 45 apically or basolaterally positive cells were identified. The efficiency of transfections was ~10%. Scale bar, 10 μ m.

METHODS. Two days after transfection, cultures were processed for immunocytochemistry by sequential staining using rabbit polyclonal subunit-specific antibodies against the $\alpha 1$ and $\beta 1$ subunits. The two GABA_A receptor-subunit-specific antibodies were generated against peptides corresponding to the C-terminal sequence of the $\alpha 1$ subunit and to the N terminus of the $\beta 1$ polypeptide. These antibodies were characterized by immunoblotting and immunohistochemistry on brain sections and neuronal cultures²⁶. For immunocytochemistry, MDCK cells were plated at a density of 2.5×10^5 cells per 35-mm dish on polylysine-coated coverslips 3½ days before transfection. At 48 to 60 h post-transfection, MDCK cells grown on coverslips were rinsed in PBS and fixed with 4% paraformaldehyde in PBS for 5 min. After washing twice with PBS, cells were permeabilized with 0.25% Triton X-100 in PBS for 10 min. Cultures were then washed three times with PBS, incubated for 10 min in 10% heat-inactivated goat serum, then washed several times in PBS. Coverslips were incubated overnight at 4 °C with anti- $\alpha 1$ polyclonal antibody at a 1:50 dilution in PBS/0.1% Triton X-100/0.1% goat serum. Coverslips were washed in PBS and then incubated for 1 h at room temperature with rhodamine-conjugated goat anti-rabbit IgG (Kirkegaard and Perry). Following three 5-min washes in PBS, coverslips were fixed a second time with 4% paraformal-



dehyde, washed, then incubated with a polyclonal anti-spectrin antibody (1:40 dilution) prepared against pig brain spectrin (by E.-h. Joe) or with a monoclonal anti-gp135 antibody (1:20; gift from G. Ojakian) for 2 h at room temperature followed by fluorescein-conjugated goat anti-rabbit or anti-mouse IgG (Kirkegaard and Perry). After incubation for 30 min at room temperature, coverslips were washed in PBS and mounted in Mowiol with 2.5% DABCO (Aldrich). Controls with secondary antibody and mock-transfected cells showed no staining. Immunofluorescence images were obtained with a Bio-Rad MRC-500 confocal laser scanning microscope fitted to a Nikon microscope in dual-channel mode using a 63 $\times$ (NA 1.3) oil-immersion lens; images were stored on an optical disk recorder and micrographs printed on a Sony videoprinter.

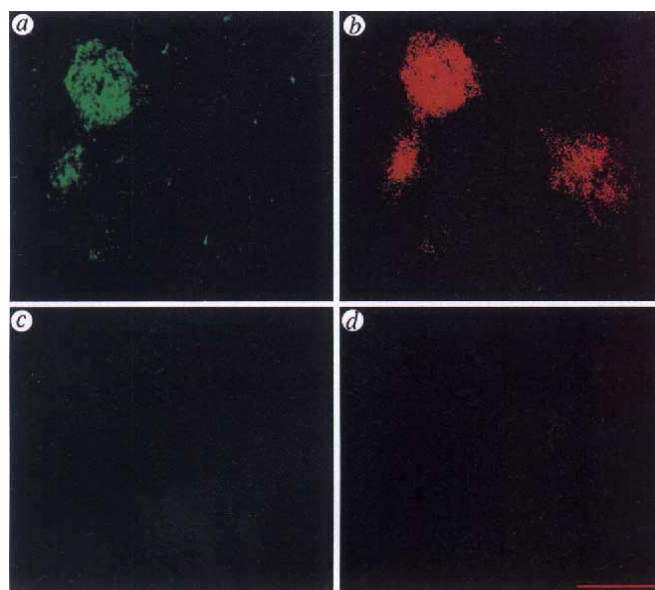


FIG. 3 Assembly of $\alpha 1$ and $\beta 1$ subunits re-routes $\alpha 1$ to the apical membrane. Double-labelled confocal micrographs showing the localization of $\alpha 1$ and $\beta 1$ subunits on the apical membrane in co-transfected MDCK cultures. *a, b*, Two cells that stain positive for both $\alpha 1$ and $\beta 1$ subunit isoforms: *a*, fluorescein channel corresponding to the $\alpha 1$ subunit, and *b*, rhodamine signal associated with expression of the $\beta 1$ subunit. *c, d*, Optical sections taken at the basolateral surface of the same cells. More than 700 cells were examined, of which 35 showed apical $\alpha 1$ and $\beta 1$ staining. Several fields showed cells in which $\alpha 1$ and $\beta 1$ were co-expressed apically in single cells together with either $\alpha 1$ (basolateral) or $\beta 1$ (apical) positive cells. Scale bar, 10 μm .

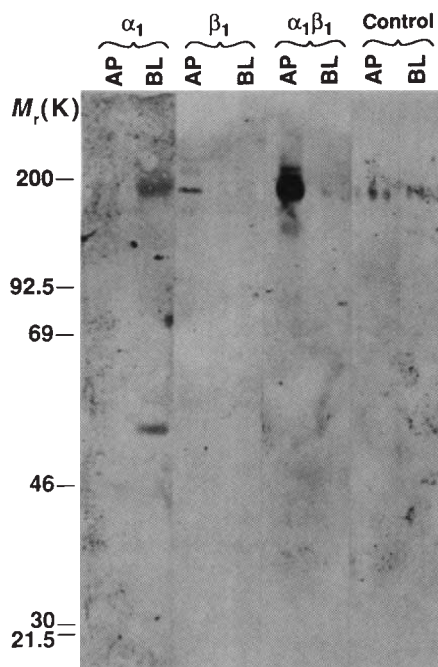


FIG. 4 Localization of $\alpha 1$, $\beta 1$ subunit proteins and the $\alpha 1/\beta 1$ protein complex on apical (AP) or basolateral (BL) membranes by biotinylation and immunoblotting. Expression of $\alpha 1$ is on the basolateral membrane ($\alpha 1$, BL) and that of $\beta 1$ is on the apical membrane ($\beta 1$, AP). Assembly of the $\alpha 1/\beta 1$ complex is shown on the apical side ($\alpha 1/\beta 1$, AP). MDCK plasma membrane proteins were labelled from the apical membrane or basolateral side and loaded onto 7.5% polyacrylamide-SDS gels and electrophoresed; $\alpha 1$: $\alpha 1$ -transfected cultures were probed with an $\alpha 1$ -specific antibody and $\beta 1$: $\beta 1$ -transfected cultures with a $\beta 1$ -specific antibody; $\alpha 1/\beta 1$ co-transfected cultures were probed with GABA_A receptor complex antibody. For the control, proteins loaded from mock-transfected cultures were probed with $\alpha 1$ - and $\beta 1$ -specific antibodies. Bands present at 200K correspond to non-dissociated complexes that have been reported for native receptors and arise because of difficulties in dissociating some GABA_A receptor subunit isoform combinations in SDS^{39,40}.

METHODS. MDCK cells grown on permeable filters were labelled on either the apical or basolateral membrane by sulphosuccinimidobiotin as described in ref. 41. Biotinylated proteins were extracted with 1% Triton X-100 and precipitated with streptavidin-Sepharose-6B. Precipitated proteins were separated on a 7.5% polyacrylamide-SDS gel in the discontinuous system of Laemmli⁴² without reducing agents and electrophoretically transferred to nitrocellulose membranes and processed according to ref. 43. Nitrocellulose membranes were probed with subunit-specific antibodies (1:25 dilution) and GABA_A receptor antibody (1:50 dilution) and detected by alkaline phosphatase-conjugated secondary antibodies.

specific vesicle carriers, and a second in which both apical and basolateral membrane proteins are inserted into and assembled at the basolateral membrane and the complex is re-routed to the apical surface by transcytosis^{31,33}. This latter mechanism is unlikely because we have never found β -subunits on the basolateral membrane at any time after transfection and because in neurons it would require GABA_A receptor subunits to be retrieved from dendrites which lie at substantial distances from the cell body or axon. Current evidence also indicates that assembly of multimeric ligand-gated ion channels, such as the acetylcholine and GABA_A receptors, occurs before insertion into the plasma membrane^{34,35}.

Like the T-cell antigen receptor, in which intermediates can be assembled from all subunit combinations but only certain combinations reach the cell surface³⁶, it is possible that GABA_A receptor complexes formed of specific subunits are architecturally edited before exit from the endoplasmic reticulum, with only specific combinations being permitted to continue to the Golgi and plasma membrane. These complexes would carry specific sorting signals for placement at particular synaptic locations, thereby conferring synapse-specific functional properties.

Whether there is a hierarchy of β -isoform-specific signals and how the neuron may accomplish their targeting to compartments is not clear.

The results reported here allow us to speculate on the possible role for the many GABA_A receptor subunits that can be expressed by an individual neuron. It has been established that there are GABAergic inhibitory inputs in the initial segment of cortical pyramidal cells and presynaptic terminals⁴⁻⁶, as well as on the somatic and dendritic membranes. Therefore if different subunits have distinct routing properties, then the placement of receptors at specific locations would depend, at least in part, on the subunit composition of the complex. In this way the plastic properties of neurons may be increased considerably because the expression of various subunit combinations, or expression of various isoforms of each subunit, may be increased by several factors including developmental differentiation^{14,15} and the excitatory synaptic drive that increases the expression of transcription factors³⁷. The up- or downregulation of different subunits and the placement and location of new complexes at new locations on the nerve cell surface, each with distinct functional properties, could be an important factor in

contributing to the plasticity of the neuron and in determining the delicate balance between excitatory and inhibitory inputs in neuronal circuits. □

Received 27 August; accepted 7 December 1992.

- Somogyi, P. et al. *J. Neurosci.* **9**, 2197–2209 (1989).
- Houser, C. R. et al. *J. Neurosci.* **8**, 1370–1376 (1988).
- Perez-Velazquez, J. L. et al. *J. Neurosci.* **9**, 2163–2169 (1989).
- Ayoub, G. S. & Matthews, G. *NeuroReport* **2**, 809–811 (1991).
- Schoch, P. et al. in *GABAergic Transmission and Anxiety* (eds Biggio, G. & Costa, E.) 11–19 (Raven, New York, 1986).
- Marin Padilla, M. *J. cogn. Neurosci.* **2**, 180–194 (1990).
- Koch, C., Poggio, T. & Torre, V. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2799–2803 (1983).
- Vu, E. T. & Krasne, F. B. *Science* **255**, 1710–1712 (1992).
- Levitani, E. S. et al. *Nature* **335**, 76–79 (1988).
- Ymer, S. et al. *EMBO J.* **8**, 1665–1670 (1989).
- Wisden, W. et al. *Neuroscience* **12**, 1040–1062 (1992).
- Gambarana, C., Pittman, R. & Siegel, E. R. *J. Neurobiol.* **21**, 1169–1179 (1990).
- Laurie, D. J. et al. *J. Neurosci.* **12**, 1063–1076 (1992).
- Killisch, I. et al. *Neuron* **7**, 927–936 (1991).
- Laurie, D. J., Wisden, W. & Seeburg, P. H. *J. Neurosci.* **12**, 4151–4173 (1992).
- Blair, L. A. C., Levitan, E. S., Marshall, J., Dionne, V. E. & Barnard, E. A. *Science* **242**, 577–579 (1988).
- Pritchett, D. B. et al. *Science* **242**, 1306–1308 (1988).
- Pritchett, D. B. et al. *Nature* **338**, 582–585 (1989).
- Verdoorn, T. A. et al. *Neuron* **4**, 919–928 (1990).
- Dotti, C. G., Parton, R. G. & Simons, K. *Neuron* **349**, 158–161 (1991).
- Dotti, D. G. & Simons, K. *Cell* **62**, 63–72 (1990).
- Powell, S. K. et al. *Nature* **353**, 76–77 (1991).
- Simons, K. & Wadinger-Ness, A. *Cell* **62**, 207–217 (1990).
- Nelson, W. J. & Veshnock, P. J. *J. Cell Biol.* **103**, 1751–1765 (1986).
- Herz, R. H. & Ojakian, G. K. *J. biol. Chem.* **264**, 4605–4612 (1989).
- Gu, Q. et al. *NeuroReport* **3**, 169–179 (1992).
- Monyer, A. et al. *Science* **256**, 1217–1221 (1992).
- Boulter, J. et al. *Science* **249**, 1033–1035 (1990).
- Rodriguez-Boulton, E. & Nelson, W. J. *Science* **245**, 718–723 (1989).
- Simons, K. & Fuller, S. D. A. *Rev. Cell Biol.* **1**, 243–268 (1985).
- Hubbard, A. L., Steiger, B. & Bartles, J. R. A. *Rev. Physiol.* **51**, 755–768 (1989).
- Wadinger-Ness, A. et al. *J. Cell Biol.* **111**, 987–994 (1990).
- Bartles, J. R. et al. *J. Cell Biol.* **105**, 1241–1248 (1987).
- Merlie, J. P. & Smith, M. M. *J. Membr. Biol.* **91**, 1–10 (1986).
- Verrall, S. & Hall, Z. W. *Cell* **68**, 23–31 (1992).
- Klausner, R. D. *New Biologist* **1**, 3–8 (1989).
- Cole, A. J. *Nature* **340**, 474–476 (1989).
- Fegner, P. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7413–7417 (1987).
- Stephenson, F. A., Duggan, M. J. & Pollard, S. J. *biol. Chem.* **265**, 21160–21165 (1990).
- Benke, D. et al. *J. biol. Chem.* **266**, 4478–4483 (1991).
- Sargiacomo, M. et al. *J. Membr. Biol.* **107**, 277–286 (1989).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Towbin, R. et al. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).

ACKNOWLEDGEMENTS. We thank S. Scherer and N. Peterson for discussion and for critically reading the manuscript. This work was supported by grants from the NIH to K.J.A. J.L.V. was an MP & B student.

DSS4-1 is a dominant suppressor of *sec4-8* that encodes a nucleotide exchange protein that aids Sec4p function

Mary Moya*, Denise Roberts* & Peter Novick†

Department of Cell Biology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

THE protein Sec4p plays an essential role at the final stage of the yeast secretory pathway and belongs to the *ras* superfamily of GTP-binding proteins¹, more specifically to a branch that includes Ypt1p in *Saccharomyces cerevisiae* and *rab* proteins in mammalian cells. GTP-binding proteins change conformation depending on whether GTP or GDP is bound² and can thus act as a regulatory switch. The protein remains in its inactive, GDP-bound form until exchange of GTP for GDP allows it to stimulate a downstream effector. This interaction is curtailed by GTP hydrolysis. The rates of nucleotide exchange and GTP hydrolysis can be regulated by interaction with accessory proteins³. Although GDP dissociation stimulators (GDS) have been identified that act on members of the *ras* and *rho* branches of the superfamily, less is known regard-

TABLE 1 Effect of yeast cell fractions or protein fractions from bacterially expressed Dss4 and Dss4-1 purification on the release of [³H] GDP from various small GTPases

Protein fractions	Substrate	Rate (min ⁻¹)	Per cent of control
BSA or buffer	Sec4	0.115	
NY13, no plasmid, S100	(yeast)	0.109	100
NY13, <i>dss4</i> plasmid, S100		0.134	123
NY13, no plasmid, P100		0.145	100
NY13, <i>dss4</i> plasmid, P100		0.208	144
NY405, no plasmid, S100		0.132	100
NY405, <i>DSS4-1</i> plasmid, S100		0.230	174
NY405, no plasmid, P100		0.162	100
NY405, <i>DSS4-1</i> plasmid, P100		0.280	173
Control fractions	Sec4	0.174	100
Dss4 protein	(Yeast)	0.618	355
Dss4-1 protein		0.617	355
Control fractions	Sec4	0.150	100
Dss4 protein	(bacterial)	0.507	338
Dss4-1 protein		0.473	315
Control fractions	<i>rab3a</i> *	0.0116	100
Dss4 protein	(bacterial)	0.0131	113
Dss4-1 protein		0.0136	117
Control fractions	Ras2	0.0107	100
Dss4 protein	(bacterial)	0.0099	93
Dss4-1 protein		0.0092	86
Control fractions	Ypt1*	0.0142	100
Dss4 protein	(bacterial)	0.0355	250
Dss4-1 protein		0.0321	230

GDP dissociation rate was determined as described in Fig. 3. In the top portion of the table, mutant and wild-type alleles were over-expressed on a 2 μ plasmid in the yeast strains, NY405 and NY13, respectively. Cultures were osmotically lysed according to Bowser and Novick²³. Supernatant from a 450g spin was centrifuged at 100,000g for 1 h at 4 °C, resulting in pellet (P100) and supernatant (S100) fractions. Fractions from strains which over-expressed Dss4p were assessed at a final concentration of 7 mg ml⁻¹. The supernatant, from strains which over-expressed Dss4-1p, was at a final concentration of 13.6 mg ml⁻¹. The corresponding pellet was at 12 mg ml⁻¹. Control fractions were at the same final concentrations as the experimental fractions. Final buffer in the presence of yeast lysates was 40 mM HEPES, pH 8.0, 100 mM NaCl, 0.5 mM GDP, 1 mM EDTA, 5 mM MgCl₂, 380 mM Sorbitol, 500 μ M triethanolamine, 1 mM DTT, 0.05% Lubrol, 1 mM phenylmethylsulphonyl fluoride (PMSF), and 1 μ g ml⁻¹ each of antipain, leupeptin, chymostatin, pepstatin and aprotinin. The presence of sorbitol in the lysis buffer slowed the intrinsic dissociation rate from 0.21 min⁻¹ to 0.12 min⁻¹. In the lower portion of the table, bacterial Sec4 was expressed and purified as in ref. 24. Bacterial Ypt1 was made and purified as in ref. 25. Experiments represent the mean of duplicate determinations from two independent experiments except for Ypt1 and *rab3a* which are representative of three similar experiments. Amounts of Dss4 (6 μ g ml⁻¹) and Dss4-1 protein (1.78 μ g ml⁻¹) were the same in the assays for yeast and bacterial Sec4 as well as for Ras2. No effect on the rate of GDP release from Ras2 was seen when Dss4 and Dss4-1 protein concentrations were increased such that acceleration of the dissociation rate from Sec4 was sixfold.

* The concentrations of Dss4 and Dss4-1 proteins used in the experiments with Ypt1 caused a fivefold acceleration in the rate of GDP release from Sec4 (data not shown) and with Rab3a caused a sixfold acceleration in the dissociation rate by Sec4 (data not shown).

ing GDSs that act on members of the Sec4/Ypt1/Rab subgroup. A preliminary characterization of a Rab3A GDP dissociation stimulating activity has been presented⁴. We report here the use of suppressor analysis to clone a gene, *dss4*, encoding a 17K protein that aids Sec4p action *in vivo* by functioning as a GDP dissociation stimulator.

An extragenic, dominant suppressor of the temperature-sensitive, *sec4-8* mutation was identified by selecting spontaneous revertants of a *sec4-8/sec4-8* diploid strain (Fig. 1). The suppressing allele has been designated *DSS4-1*, to indicate that it is a dominant suppressor of *sec4*. This allele allows growth of *sec4-8* cells up to 35 °C, but it does not suppress a deletion of *SEC4*. *DSS4-1* does not suppress the growth defects of other,

* M.M. and D.R. are equal first authors.

† To whom correspondence should be addressed.

post-Golgi blocked *sec* mutants, nor does it suppress the growth defect of a *yp1-1* mutant (not shown). The effect of *DSS4-1* on the efficiency of invertase secretion by *sec4-8* and *SEC4* cells was tested as previously described⁵. Wild-type cells secrete 97% of the invertase synthesized during a 1 h incubation at 34 °C. A *sec4-8* strain secretes only 24%, whereas a *sec4-8, DSS4-1* strain secretes 64%. A *SEC4, DSS4-1* strain secretes 93% of the invertase, consistent with the observation that *DSS4-1* has no effect on growth in a *SEC4* background. The *sec4-8* secretory block leads to the accumulation of secretory vesicles at elevated temperatures⁶. *DSS4-1* dramatically reduces the number of vesicles seen at 34 °C in a *sec4-8* strain (not shown). A *SEC4, DSS4-1* strain does not accumulate vesicles. Thus, the *DSS4-1* mutation suppresses the secretion defect and vesicle accumulation

phenotype of *sec4-8*, and, furthermore, does not lower the efficiency of secretion or lead to vesicle accumulation in a *SEC4* background.

DSS4-1 was cloned by its ability to suppress the growth defect of a *sec4-8* strain (Fig. 1a). The sequence contains an open reading frame of 429 nucleotides encoding a protein of 143 amino acids that has an *M_r* of 17K and is generally hydrophilic, lacking a hydrophobic domain long enough to span the lipid bilayer. We screened the NBRF library for similar sequences, but found none homologous to the *Dss4-1* protein sequence. The wild-type *dss4* allele was isolated by screening a genomic yeast library in phage λ EMBL3A, using as a probe the 1.1 kilobase (kb) *Hind*III fragment of pNB526. Comparison of the two sequences showed that the suppressing allele contained a

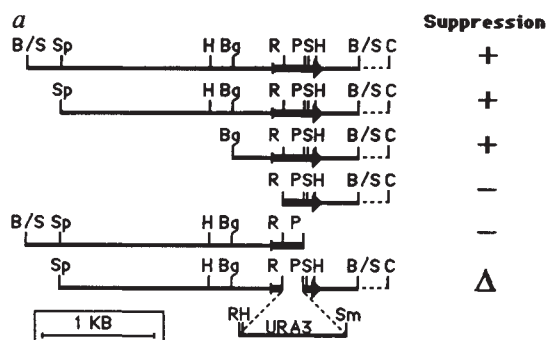


FIG. 1 a, Restriction map of pNB526 which was isolated from a genomic library by its ability to suppress the growth defect of a *sec4-8* strain at 33 °C. The library was derived from NY511 (*sec4-8, DSS4-1*) in the centromere-based shuttle vector, YCp50. Analysis of subclones demonstrated that the suppressing activity was localized to the right of the *Bgl*II site and that the region between *Eco*RI and *Pvu*II is completely internal to *DSS4-1*. This was deleted and replaced with an *Eco*RI–*Sma*I fragment carrying the *URA3* gene in order to construct a *dss4Δ* allele. Restriction enzyme sites: B/S, *Bam*HI–*Sau*IIa junction; Sp, *Sph*I; H, *Hind*III; Bg, *Bgl*II; R, *Eco*RI; P, *Pvu*II; S, *Ssp*I; C, *Cla*I. The heavy arrow shows the open reading frame of *dss4*; the dashed line indicates vector sequence. b, The DNA sequence of the *dss4* locus, and the amino-acid sequence of the *Dss4* protein. The nucleic acid sequence of *DSS4-1* was identical to *dss4* except for the replacement of the G at position 323 in *dss4* with an A in *DSS4-1*.

METHODS. *DSS4-1* was identified by selecting spontaneous revertants of a homozygous diploid strain NY443 (*MATa/MATα, sec4-8/sec4-8, his4-619/+ , ura3-52/+*) at 33 °C. Analysis of tetrads derived from one such revertant demonstrated 2:2 growth at 33 °C, indicative of a single nuclear mutation that confers temperature resistance. A cross of one of the temperature-resistant haploid products to a wild-type strain demonstrated that the suppressor was unlinked to the *sec4* locus. Both loci are, however, centromere linked: of 51 tetrads analysed, 16 were parental ditype, 31 were nonparental ditype and only 4 were tetratype. Because the *DSS4-1* mutation does not confer any recessive properties that would aid cloning of the wild-type allele, it was necessary to prepare a genomic plasmid library from a *sec4-8, DSS4-1* strain and to clone the mutant allele by its ability to suppress the growth defect of a *sec4-8* strain. Chromosomal DNA (100 μg), isolated from NY511 (*sec4-8, DSS4-1*) according to ref. 10, was digested with 6 units *Sau*IIa and electrophoresed on 0.6% agarose gel. The 3–10 kb region was excised and DNA was extracted. One third of the digested DNA was ligated with 4 μg YCp50 DNA which had been digested with *Bam*HI and treated with phosphatase. The ligation mix was introduced into DH1 cells and 40,000 transformants were obtained. Plasmid DNA was prepared from an amplified pool of transformants and introduced into NY405 (*ura3-52, sec4-8*). After selection on SD plates at 25 °C, 4,000 transformants were replica stamped onto YPD media and incubated at 34 °C. Plasmids were recovered from 10 temperature-resistant colonies. The recovered plasmids fell into two classes. One class had in common several restriction fragments indicative of the *sec4-8* gene. This was expected because multiple copies of *sec4-8* permit growth at elevated temperatures¹. The remaining three plasmids did not contain the *sec4* gene, but each contained a 1.06 kb *Hind*III fragment, suggesting that they were structurally related. The smallest, pNB526, was restriction mapped more extensively. Subclones were prepared in YCp50 and introduced into NY405 (*ura3-52, sec4-8*). After selection on

b			
-371	GATCTTTCTTTTTCAGTTGAAGCCTTTCACGCGTCCGAACGCTAGCATATATGCAAGC	-312	
-311	CGTCTGACACTATAGATCGCGCCAGTAGGGAACCACTTGATCCAGCAGATGAGCCGA	-252	
-251	CAACTCAAAACCTCCCTTCTCTGTATCAAAATACCTTTTCAGCATATCCATATCATTCTT	-192	
-191	CCTTTCTCTTCGTAAGGCTGACTGGGCGCCGGGACCCGCTAGTTTAGTAGTACTGTGA	-132	
-131	ATATCGTCATTGTCTACTGAACGATTATATATCGATCATTTTAAATGAAGTACATA	-72	
-71	TCAAAACCTTGACTGGTATGATAGAGAAAGTTCAATTAAGGGCTGGAAGAAATGGGGAA	-12	
-11	TAACAGAGAACATGAGCAAGGCTACGTGTTCTTTGAGGGCTGTCTAGTGCAGTTATTA	48	
	M S K A T C S F E G C H S A V I T		
49	CCATCAACGATGATAACATAAATACCTTACCAGAGCAGGTGCACGCGAATCAAGTTGC	108	
	I N D D N I I C A D D K G P I G I C K L		
109	TGGAAACAGAACTATGCGAGATGCTACCCCGAGGAGAGCACTTCTTGGTCTGCTCCCTG	168	
	E N R T T M R D A T P S E S N F L V V P D		
169	ATGTTTGGGACTTTGATAATGTTGGTGTATCTAGAGAAATCCCTTCTCTATATTAGGGG	228	
	V W D F D N V G V S R E I P S S I L G D		
229	ATTGAGTGATAAGAGTGATTTTGTCTTGTAGTATGTTTCCAGCTGGAAGATTAATA	288	
	L S D K S D F V F E Y G N S S W K I K K		
289	AGTGTGAAATATTTGATATGCGCTGATGTTGATAAGGCTCCTATGGAATAATATGTA	348	
	C L K Y L I C A D C D K G P I G I C K L		
349	AGTGCAAGATCAGACAAAGAAATGAAGAAAGAGTTTGCATCTACTAAGCTTGCAGACGC	408	
	V Q D Q T K N E E R V L H L L S L R T L		
409	TTCAATATGGAAGAAATAGTTTGGTAGGCACTAGACAAAATAGTTGCCGCCAC	468	
	Q I M G R N		
469	CAGAAAAAAGAGTAACCAACCACTATTTCCACTGGCACTGGGCCCTATACCA	528	
529	GCTATAGAGTAGTCTCTGTTTACATTACTAGCATCACTTTATGTAATAATATATGGCAT	588	
589	TCTGGACGAATCCTAATGGTAGACATTACCCGTTTTTCGAGGGCCACAAATTCGGAA	648	
649	ATTTTATTAATT	661	

SD plates at 25 °C, the *Ura*⁺ transformants were tested for growth on YPD plates at 34 °C. To verify that pNB526 carried *DSS4-1*, we determined whether the cloned sequence could direct integration to the genomic *dss4* locus. The fragment was subcloned into the vector Yip5, which carries the *URA3* gene, but no yeast origin of replication. After cleavage at a unique *Bgl*II site in the insert, the plasmid was used to transform a *sec4-8, ura3-52* strain to *Ura*⁺. This results in the introduction of the *URA3* gene between two copies of the cloned sequence and in suppression of the temperature sensitivity. The transformants were crossed to a *sec4-8, DSS4-1* strain, and tetrads were analysed. Because all progeny grew at 34 °C, we conclude that the site of integration of the cloned suppressor was tightly linked to the *dss4* locus. Together these data establish that the cloned sequence contains the *DSS4-1* gene. The *dss4* allele was isolated by screening a wild-type yeast genomic library inserted in phage λ EMBL3A according to ref. 11 using as a probe the ³²P-labelled 1.06 kb *Hind*III fragment isolated from pNB526. Host bacteria used during the screening was *E. coli* strain Y1090. Prehybridization, hybridization, washing of the blots and preparation of the probe were as in ref. 12. Phage DNA purified from positive plaques was digested with *Dra*I. A 1.8 kb *Dra*I fragment carrying the *dss4* region was inserted into the *Eco*RV site of the pBlueScript SK +/– to obtain pNB475. The multi-copy plasmid carrying the wild-type gene *dss4* was obtained by insertion of the *Cla*I–*Bam*HI fragment of pNB475 into the pRS426 2μ vector¹³ that had been digested with *Cla*I and *Bam*HI. Template DNA for sequence analysis was obtained by subcloning the original 3 kb insert of pNB526 containing *DSS4-1* and the 1.8 kb insert of pNB475 containing *dss4* into pBlueScript SK +/–. Template was isolated from the *E. coli* strain TG1. The nucleotide sequence was determined as in ref. 14. The predicted protein sequence was compared with the National Biomedical Research Foundation library by the FASTP program in the KTUP1 mode¹⁵.

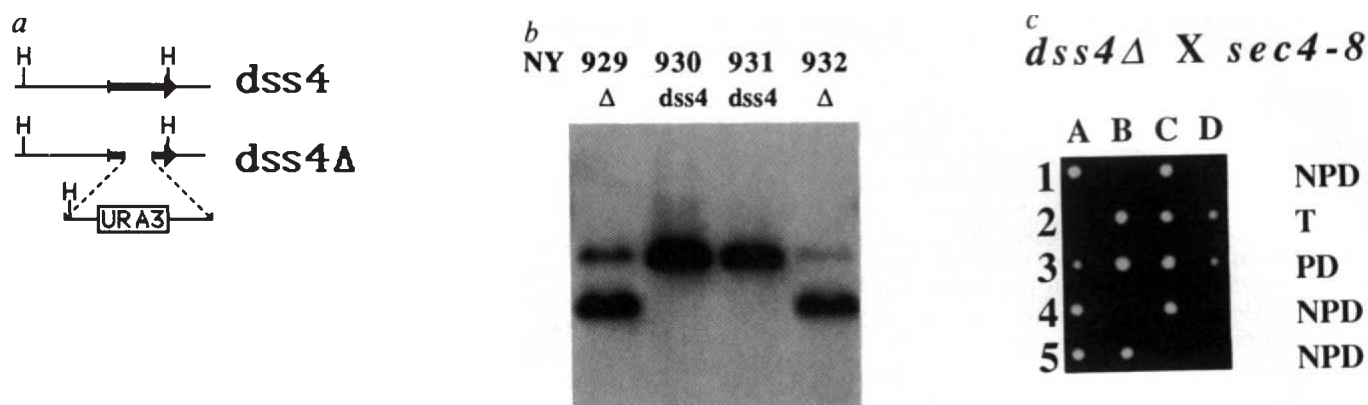


FIG. 2 *a*, Predicted structure of the *dss4* locus before and after disruption. *b*, DNA was isolated from the meiotic products of a tetrad from a *dss4/dss4Δ* diploid. DNA was digested with *Hind*III and analysed by Southern blot using the *Hind*III fragment as a probe. The *dss4* colonies show one hybridizing band at 1.06 kb, whereas *dss4Δ* colonies show two bands, as predicted. The upper band weakly hybridizes because of short complementarity with the probe. *c*, A *dss4Δ*, *SEC4* strain was crossed to a *dss4*, *sec4-8* strain and tetrads were analysed. Representative colonies are as follows: (*SEC4*, *dss4*), 1A and 1C; (*sec4-8*, *dss4Δ*), 1B and 1D; (*SEC4*, *dss4Δ*), 3B and 3C; (*sec4-8*, *dss4*), 3A and 3D. NPd, Non-parental ditype; T, tetatype; PD, parental-ditype. *d*, Lysates were prepared from *dss4*, *dss4Δ* and *DSS4-1* strains. After SDS-PAGE electrophoresis and transfer to nitrocellulose, the filter was probed with affinity-purified anti-Dss4p antibody and with 125 I-protein A. The arrow on the left indicates the Dss4 protein, and the arrow on the right indicates the Dss4-1 protein.

METHODS. Total yeast DNA, extracted as in ref. 16 and digested with *Hind*III, was fractionated by electrophoresis in 0.5% agarose gel and transferred to nitrocellulose¹⁷. Prehybridization, hybridization, washing of the blots and preparation of the probe (32 P-labelled 1.1 kb *Hind*III fragment from pNB526)

were according to ref. 12. The plasmid pNB479, containing the TrpE gene fused in-frame to an internal fragment of *DSS4-1* gene, was constructed by inserting the 0.3 kb *Hind*III-*Eco*RI fragment from pNB526 into pATH11 (ref. 18) which had been digested with *Hind*III and *Eco*RI. Fusion protein was produced in the *E. coli* strain DH1 and electroluted from SDS-PAGE¹⁹. Immunization of rabbits was as in ref. 20. The affinity-purification of antibodies was according to ref. 21.

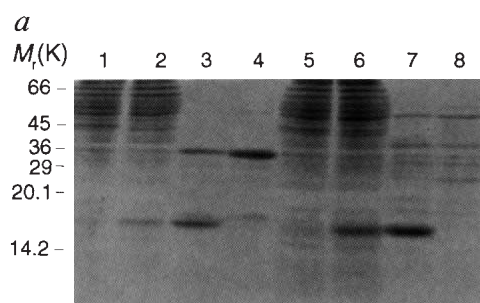
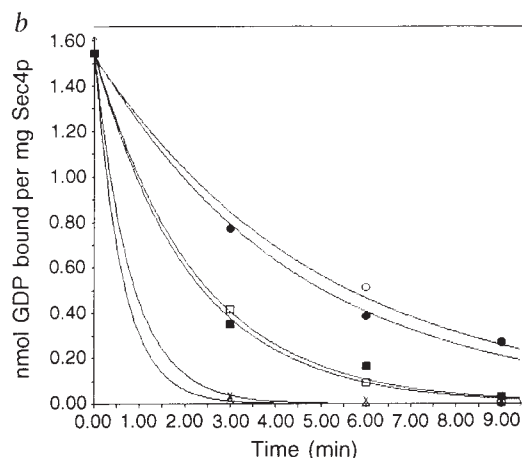


FIG. 3 *a*, Coomassie-stained, 17% SDS polyacrylamide gel showing partial purification of bacterially produced Dss4-1 and Dss4 proteins by gel filtration chromatography. Low-speed supernatant after lysis of bacteria containing expression vector alone (lanes 1 and 5); low-speed supernatant from lysed bacteria expressing Dss4p (lane 2); pooled gel filtration fractions containing Dss4p (lane 3); corresponding pooled fractions from gel filtration of vector-only lysates (lane 4); low-speed supernatant from lysed bacteria expressing Dss4-1p (lane 6); pooled fractions containing Dss4-1p after gel filtration (lane 7); corresponding pooled fractions after gel filtration of vector-only lysates (lane 8). M_r markers are indicated on the left. *b*, Recombinant Dss4 and Dss4-1 proteins, purified gel filtration chromatography, accelerated the GDP dissociation rate of Sec4p. Dss4p was at a final concentration of $3 \mu\text{g ml}^{-1}$ (■) and $18 \mu\text{g ml}^{-1}$ (Δ), Dss4-1p was at a final concentration of $0.89 \mu\text{g ml}^{-1}$ (□) and $5.34 \mu\text{g ml}^{-1}$ (X), Vector-only, pooled, S100 fractions corresponding to pooled fractions containing Dss4p (●) or Dss4-1p (○) were at a final concentration of $3 \mu\text{g ml}^{-1}$ and $5.34 \mu\text{g ml}^{-1}$, respectively. Data shown represent the mean of duplicate determinations from two independent experiments.

METHODS. The T7 expression system⁸ was used to express native Dss4 and Dss4-1 protein in *E. coli* strain BL21(DE3) using the expression vector, pET11d. The *dss4* and *DSS4-1* coding sequences were amplified using polymerase chain reaction (PCR) and ligated behind the T7 promoter. After induction of protein expression with 1 mM isopropyl- β -D-thiogalactoside (IPTG) for 1 h at 37 °C, pelleted cells were resuspended in HEPES, pH 8.0 buffer (20 mM HEPES, 100 mM NaCl, 5 mM MgCl_2 , 1 mM DTT, 1 mM PMSF



and $1 \mu\text{g ml}^{-1}$ each of antipain, leupeptin, chymostatin, pepstatin and aprotinin.) Cells were sonicated on ice and centrifuged at low speed (10,000g for 15 min) at 4 °C. Dss4-1 and Dss4 proteins were purified by fractionation through a Sephacryl S100 column equilibrated in 50 mM HEPES, 100 mM NaCl, 5 mM MgCl_2 , 1 mM DTT (175 ml column, 2.2 ml fractions). The presence of Dss4 and Dss4-1 proteins in fractions were assessed by immunoblot. Pooled peak fractions were analysed by SDS-PAGE and densitometric analysis to quantify the amount of Dss4p and Dss4-1p in their respective preparations. The GDP dissociation rate in the presence of partially purified, recombinant Dss4 and Dss4-1 protein was a modification of ref. 7. Briefly, 100 mM Sec4 protein was incubated with $0.2 \mu\text{M}$ [$8, 5$ - ^3H]-GDP (Du Pont-New England Nuclear), at a specific activity of 20,000 c.p.m. pmol^{-1} at 30 °C. Cell fractions, supplemented with 1 mM cold GDP, were added after 60 min. Final buffer was 50 mM HEPES, pH 8.0, 120 mM NaCl, 0.5 mM GDP, 1 mM EDTA, 6 mM MgCl_2 , 1 mM DTT, 0.05% Lubrol. Sec4p was at a final concentration of 25 nM. Aliquots of the reaction mixture were removed over time and filtered through Millipore, nitrocellulose filters. Protein-associated radioactivity was assessed in a scintillation counter. Data were fitted to an exponential decay equation using non-linear regression analysis²² using a computer program, Enzfitter. Yeast Sec4p was purified as in ref. 7.

single A to G mutation at position 323 which led to a change in amino acid 108 from aspartate in the wild-type gene to glycine in the suppressing allele (Fig. 1b). This amino-acid change is precisely complementary to the *sec4-8* mutation which changes the glycine at position 147 in wild-type Sec4p to an aspartate¹.

To determine if the *dss4* gene is essential, the *URA3* gene was used to replace the internal *PvuII-EcoRI* fragment of one chromosomal copy of *dss4* in a diploid strain, homozygous for the *ura3-52* mutation (Fig. 2a). The *Ura*⁺ diploid transformants were shifted to sporulation media, and tetrads were dissected. All four spores of each tetrad were viable, and two were *Ura*⁺. To verify the correct integration of the disrupted gene at the chromosomal locus, we analysed the genomic DNA extracted from four spores of the same tetrad, NY929-932, by Southern blot. The two *Ura*⁺ spores carrying the disrupted gene showed the two bands predicted by the restriction map (Fig. 2b), whereas the two *Ura*⁻ spores which carried the non-disrupted gene showed only one band. Thus, the *dss4* gene is not essential.

As disruption of *dss4* was not lethal, we crossed the *dss4Δ* strain, NY932, with all the *sec* mutants to detect possible synthetic effects. The diploids were shifted to sporulation media, and tetrads were analysed. A strong negative effect on growth at 25 °C was detected in the double mutant spores carrying the disruption of *dss4* gene in combination with the mutant alleles *sec2-49*, *sec4-8*, *sec8-9* and *sec15-1* (Fig. 2c). A weaker effect on growth at 35 °C could be shown with the *sec3-2* mutant. This subset of *sec* mutants (*sec2-49*, *sec8-9* and *sec15-1*) can be suppressed by a duplication of the *SEC4* gene and shows lethality in combination with the *sec4-8* mutation. Because these results show that the loss of *dss4* can aggravate as well as mimic a partial loss of *SEC4*, we suggest that the normal role of Dss4p is to aid Sec4p function.

Polyclonal rabbit antibody was raised against the central portion of Dss4p (amino acids 33 to 133) and affinity purified. Immunoblot analysis of a wild-type lysate demonstrated two reactive bands close to the predicted *M_r* of 17K. The lower band of the doublet was shifted to a slightly faster mobility in a lysate of *DSS4-1* cells and was absent in a lysate of *dss4Δ* cells (Fig. 2d), establishing it as the Dss4 protein. The upper band is seen in lysates derived from *dss4*, *DSS4-1* as well as *dss4Δ* strains and, therefore, represents a crossreacting protein.

The *dss4* and *DSS4-1* genes were inserted into multi-copy number yeast vectors and introduced into *SEC4* and *sec4-8* strains. Overexpression of *DSS4-1* did not interfere with growth of either strain at 25 °C and, as expected, restored the growth of *sec4-8* cells at 34 °C. Over-expression of the *dss4* gene in *sec4-8* cells did not restore growth at 34 °C and, moreover, severely impaired growth at 25 °C. Growth of *SEC4* cells was unaffected by overexpression of *dss4*.

The strong genetic interactions of *DSS4-1*, *dss4Δ* and *dss4* with *sec4-8* suggested the possibility of a direct association of the gene products. Exchange proteins, which alter the rate of guanine nucleotide dissociation, are among the proteins that physically interact with members of the *ras* superfamily³. To determine if Dss4p functions as an exchange protein, membrane

and soluble fractions derived from yeast over-expressing either *dss4* or *DSS4-1* were compared to fractions derived from control strains with regard to their effect on the rate of dissociation of GDP from Sec4p.

GDP dissociates from Sec4p with a rate of 0.21 min⁻¹ at 30 °C (ref. 7 and Fig. 3). Addition of a 100,000g pellet fraction, derived from a yeast strain in which the *dss4* gene was over-expressed, accelerated the GDP dissociation rate of Sec4p by 1.5-fold compared to the control pellet which was generated from isogenic strains lacking the multi-copy number plasmid (Table 1). This release-stimulating activity was not associated with the 100,000g supernatant. Immunoblot analysis confirmed the presence of Dss4p in the pellet, but not in the supernatant (data not shown). Both the 100,000g pellet and supernatant fractions derived from strains over-expressing *Dss4-1* accelerated the release of GDP from Sec4p up to 1.8-fold compared to control fractions from isogenic strains without the over-expressing plasmid. The over-expressed Dss4-1 protein was found in both the pellet and supernatant by immunoblot analysis (data not shown).

To determine if partial purification would enable detection of a more dramatic effect on the rate of Sec4p nucleotide release, Dss4-1 and Dss4 proteins were expressed in *Escherichia coli* by the method of Studier *et al.*⁸, and cytosolic fractions were chromatographed by gel filtration on a Sephacryl S100 column (Fig. 3a). Pooled column fractions containing Dss4 or Dss4-1 proteins, as determined by immunoblot, (data not shown) accelerated the rate of release of GDP up to six- and eight-fold, respectively, compared to the intrinsic dissociation rate (Fig. 3b). Equivalent column fractions derived from a strain carrying the expression vector alone had no effect on the intrinsic rate. Dss4 and Dss4-1 proteins also accelerated the rate of GDP release from Sec4 which had been purified from bacteria, suggesting that prenylation is not absolutely required for Dss4 to act on its substrate (Table 1).

We examined the interaction of Dss4 with other small GTPases using the *in vitro* filter assay (Table 1). Dss4 and Dss4-1 proteins did not significantly alter the rate of GDP release from bacterially expressed Rab3a or Ras2. Ypt1 is more closely related to Sec4, sharing 47.5% identity¹. GDP release from bacterial Ypt1 was accelerated by both Dss4 and Dss4-1 proteins, but the effect was not as strong as in the case of Sec4. Amounts of Dss4 and Dss4-1 protein that caused a fivefold acceleration in the rate of GDP release by bacterial Sec4 resulted in only a 2.5-fold acceleration of the GDP dissociation rate from Ypt1 (Table 1).

The genetic and biochemical evidence indicates that *dss4* encodes a protein that aids Sec4p function in the cell by catalysing guanine-nucleotide exchange. The Dss4 protein does not have sequence similarity with known exchange proteins. Significant similarity, however, has been observed with the mammalian suppressor of *sec4-8*, designated *mss4* (ref. 9). Dss4p and Mss4 may represent a family of exchange proteins that acts on members of the Sec4/Ypt1Rab branch of the Ras superfamily. □

Received 20 July; accepted 27 November 1992.

- Salminen, A. & Novick, P. J. *Cell* **49**, 527-538 (1987).
- Milburn, M. V. *et al.* *Science* **247**, 939-945 (1990).
- Bollag, G. & McCormick, F. A. *Rev. Cell Biol.* **7**, 601-632 (1991).
- Burnstein, E. S. & Macara, I. G. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1154-1158 (1992).
- Nair, J., Müller, H., Peterson, M. & Novick, P. J. *Cell Biol.* **110**, 1897-1909 (1990).
- Novick, P. J., Field, C. & Schekman, R. *Cell* **21**, 205-215 (1980).
- Kabcenell, A. K., Goud, B., Northup, J. & Novick, P. J. *J. Biol. Chem.* **265**, 9366-9372 (1990).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorf, J. W. *Meth. Enzym.* **185**, 60-89 (1990).
- Burton, J., Roberts, D., Montaldi, M., Novick, P. & De Camilli, P. *Nature* (in the press).
- Philippson, P., Stoltz, A. & Scherf, C. *Meth. Enzym.* **194**, 169-181 (1991).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
- Emanuel, J. T. *et al.* *Molec. cell. Biol.* **6**, 2476-2481 (1986).
- Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. *Gene* **110**, 119-122 (1992).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).

- Lipman, D. & Pearson, W. *Science* **227**, 1435-1441 (1985).
- Holm, C., Meeks-Wagner, D., Fangman, W. & Botstein, D. *Gene* **42**, 169-173 (1986).
- Southern, E. M. *J. molec. Biol.* **98**, 503-518 (1975).
- Koerner, T. J., Hill, J. E., Myers, A. M. & Tzagoloff, A. *Meth. Enzym.* **194**, 477-490 (1991).
- Hurkapiller, M., Lujan, E., Oitrand, F. & Hood, L. *Meth. Enzym.* **91**, 227-236 (1983).
- Goud, B., Salminen, A., Walworth, N. C. & Novick, P. J. *Cell* **53**, 753-768 (1988).
- Guesdon, J. & Avrameas, S. *J. Immun. Meth.* **11**, 129-133 (1976).
- Marquardt, D. W. *J. Soc. ind. appl. Math.* **11**, 431-441 (1963).
- Bowser, R. & Novick, P. J. *J. Cell Biol.* **112**, 1117-1131 (1991).
- Novick, P., Garrett, M., Brennwald, P. & Kabcenell, A. K. *Meth. Enzym.* **219**, 352-362 (1992).
- Walworth, N., Brennwald, P., Kabcenell, A., Garrett, M. & Novick, P. *Molec. cell. Biol.* **12**, 2017-2028 (1992).

ACKNOWLEDGEMENTS. We thank M. Synder in the Department of Biology at Yale University for the EMBL3A genomic library, T. Sudhof for Rab3a, S. Power for Ras2 and M. Garrett for preparation of bacterial Sec4 and Ypt1. This work was supported by grants to P.N. from the NIH. M.M. was supported by the Mathers foundation and D.R. was supported by the NIH.

A mammalian guanine-nucleotide-releasing protein enhances function of yeast secretory protein Sec4

Janet Burton^{*†}, Denise Roberts[†], Maribeth Montaldi^{*†}, Peter Novick[†] & Pietro De Camilli^{*†}

^{*}Howard Hughes Medical Institute and [†]Department of Cell Biology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

SMALL GTP-binding proteins of the *ras* superfamily are important for exocytosis from eukaryotic cells¹⁻⁵. GTP-binding proteins can exist in two different conformations depending on whether they are bound to GDP or GTP, and are thought to function as molecular switches that regulate a variety of cellular processes^{1,3}. The GTP-GDP cycle is controlled by accessory proteins that promote the exchange of bound GDP or the hydrolysis of GTP. The protein Sec4, a member of the Sec4/Ypt1/Rab branch of the Ras superfamily, is involved in a late stage of the secretory pathway in yeast⁶. Here we report the isolation of a mammalian complementary DNA, *mss4*, encoding a GDP-releasing protein that enhances Sec4 function. The Mss4 protein also stimulates GDP release from Ypt1 and from the mammalian protein Rab3a, but not from Ras2. Mss4 shows sequence similarity to Dss4, a yeast protein with similar biochemical properties.⁷

Rat brain cDNAs were cloned into an expression vector containing the yeast alcohol dehydrogenase promoter and a uracil selectable marker⁸ (Fig. 1). The cDNAs were expressed in *sec4-8* yeast cells, which contain a temperature-sensitive allele of *SEC4*⁹, and screened for suppression of the growth defect at the restrictive temperature. A temperature-resistant clone was isolated. Temperature-resistance was recovered when this plasmid was re-transformed into the *sec4-8* parent strain. The isolated cDNA was named *mss4*, for mammalian suppressor of *sec4*. At 37 °C, the *sec4-8* strain with the *mss4* cDNA-containing plasmid (JBY2) grew at a rate comparable to that of a wild-type *SEC4* strain (compare Fig. 1b with a). In contrast, the *sec4-8* strain containing the expression vector alone (JBY1) and the non-transformed *sec4-8* parent strain were unable to grow at this temperature (Fig. 1c and d).

The efficiency of invertase secretion from the *mss4*, *sec4-8* strain was tested to determine whether *mss4* restored growth to the *sec4-8* strain by suppressing the secretion defect. When shifted to the restrictive temperature for one hour in low glucose, wild-type *SEC4* cells secrete greater than 90% of the invertase produced, while strains containing the mutant *sec4-8* accumulate invertase intracellularly in post-Golgi secretory vesicles^{6,9}. In the present study under the same conditions, wild-type cells secreted 96%, *sec4-8* cells secreted 17%, the vector-alone strain secreted 19% and the *sec4-8*, *mss4* strain secreted 33% of the invertase produced. Consistent with the increase in invertase secretion, the *sec4-8*, *mss4* strain also had a reduction in the number of accumulated post-Golgi secretory vesicles at the restrictive temperature when compared to the *sec4-8* parent strain as assessed by electron microscopy (not shown). Therefore, the *mss4* cDNA restores growth of the *sec4-8* strain by suppression of the secretion defect.

The *mss4* cDNA is 1.7 kilobases (kb) in length. The nucleotide sequence was determined and found to contain an open reading frame (ORF) predicting a 123-amino-acid protein (Fig. 2a). The 800-bp *Xba*I fragment containing this region restores growth to the *sec4-8* strain; other regions of the *mss4* cDNA are unable to suppress the growth defect (Fig. 2b). If this ORF is truncated or placed in the reverse orientation with respect to the promoter, suppression is lost. However, as there were no stop codons

upstream of the predicted start methionine, it was not clear that the cDNA encoded the full-length protein. We therefore screened a λ ZAPII rat brain cDNA library with the 1.7-kb *mss4* cDNA and isolated a 2.5-kb cDNA that was identical to the original clone except that it contained an additional 740 bp of upstream sequence. Three in-frame stop codons located at positions -82, -100 and -121 bp relative to the start methionine were found (Fig. 2a). The predicted 123-amino-acid ORF, therefore, represents the full-length Mss4 protein responsible for suppression of the *sec4-8* mutation.

Northern blots were done on messenger RNAs from various rat tissues using the 1.7-kb *mss4* cDNA as probe. Two mRNA transcripts of ~2.5 and 1.5 kb were found in all of the tissues tested (Fig. 2c). These two mRNAs may represent two highly homologous genes or an alternative splicing event. As the longest cDNA clone was 2.5 kb, it should correspond to the larger of the two transcripts.

Based on the amino-acid sequence, the Mss4 protein has a relative molecular mass (M_r) of 14,000 (14K), is hydrophilic¹⁰, and contains no potential membrane-spanning domains. Mss4 showed no significant homology with any protein sequences in the Swiss protein or PIR databases. Furthermore, Mss4 contains no GTP-binding consensus sequences³ and is therefore not a

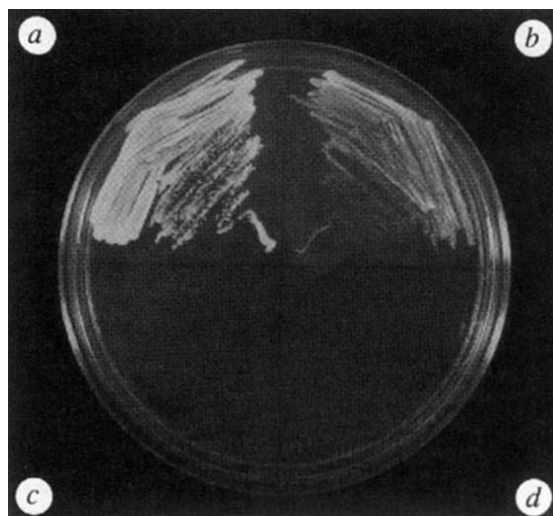


FIG. 1 The *mss4* cDNA restores growth to the *sec4-8* strain when grown at 37 °C. a, NY15, a *SEC4* wild-type yeast strain; b, JBY2, *mss4*, *sec4-8* strain; c, NY405, *sec4-8* parent strain; d, JBY1, *sec4-8* strain containing expression vector alone.

METHODS. The above strains were streaked onto a rich-medium plate (YPD) and grown for 24 h at 37 °C. Rat brain cDNAs were made from 5 μ g poly(A)⁺ RNA isolated from total brain of adult rats¹⁸. The cDNAs were ligated to kinased *Xho*I adaptors, size-selected on 1.2% agarose¹⁸, inserted in the pVT102U yeast expression vector⁸ at the *Xho*I site, then introduced into Electromax DH10B cells (BRL) using a BRL electroporation apparatus. Colonies were pooled and a plasmid DNA was prepared¹⁸. The plasmid library was introduced into the *sec4-8* strain using a high-efficiency spheroplast transformation procedure¹⁹. Ura⁺ transformants were allowed to grow for 2 days at 25 °C on selective medium (SD, without uracil); plates were then transferred to 37 °C for two days to allow temperature-resistant macrocolonies to form. A pure population of each temperature-resistant clone was isolated. All temperature-resistant clones were grown in rich medium at 25 °C, allowing loss of the cDNA-containing plasmid; clones were subsequently replica-plated onto both selective medium at 25 °C and rich medium at 37 °C. Clones unable to grow in selective medium (having lost the plasmid) but still able to grow at 37 °C were revertants of the *sec4-8* allele and were discarded. One temperature-resistant clone (JBY2) was isolated; upon loss of the plasmid, JBY2 became temperature-sensitive and uracil-dependent. The plasmid (11A) containing the cDNA from JBY2 was isolated and then re-introduced into the *sec4-8* parent strain. This confirmed that the plasmid was responsible for the temperature resistance of this strain.

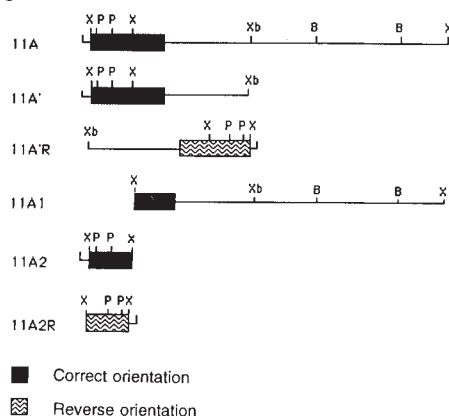
FIG. 2 Nucleotide and amino-acid sequence of *mss4* cDNA. **a**, The 2.5-kb *mss4* cDNA and predicted 123-aa protein are shown. The first nucleotide of the 1.7-kb cDNA is underlined. Both strands of the cDNA were sequenced using Sequenase (USB). The three in-frame stop codons upstream of the start methionine are underlined. Two potential protein kinase C phosphorylation sites and a potential casein kinase II phosphorylation site are marked with an open arrowhead and a filled arrowhead, respectively. **b**, Restriction map of the *mss4* cDNA in which the open reading frame encoding *Mss4* is boxed. The alcohol dehydrogenase promoter (not shown) is on the left-hand side of all constructions. Each construction's ability to rescue the growth defect of the *sec4-8* strain is shown on the right. X, *Xho*I; P, *Pst*I; Xb, *Xba*I; B, *Bgl*II. **c**, Northern blot hybridization analysis of RNA from several rat tissues using the 1.7 kb *mss4* cDNA as a probe. In lanes 1–5, 5 µg of mRNA were loaded from brain, heart, kidney, lung and liver, respectively. In lane 6, 5 µg ribosomal RNA was loaded. **METHODS.** A rat brain λZapII cDNA library (Stratagene) was screened with the 1.7-kb *mss4* cDNA as a probe using methods described previously (ref. 18). 11A is the plasmid containing the 1.7-kb *mss4* cDNA. 11A' and 11A'' were constructed by digesting 11A with *Xba*I, which cuts upstream of the insert and downstream of the open reading frame, generating an 800-bp *Xba*I fragment which was subcloned into the *Xba*I site of pVT102U¹⁸. 11A1 and 11A2 were constructed by the digestion of 11A with *Xho*I; the resulting fragments, 1.5 kb for 11A1 and 200 bp for 11A2, were each subcloned into the *Xho*I site of pVT102U. All constructs were tested for their ability to restore growth to the *sec4-8* strain; this strain was transformed using the alkali cation method²⁰ and the resulting Ura⁺ transformants were grown for 24 h at 37 °C. Total RNA was isolated from the above tissues using the guanidine isothiocyanate procedure²¹. Poly(A)⁺ RNA was isolated by oligo-dT cellulose¹⁸. RNAs were run on formaldehyde agarose gels as described previously¹⁸. The RNA was transferred to Zeta-Probe nylon membrane (Bio-Rad) 2 h at 0.8 A in 1 × TAE¹⁸. The blot was hybridized in 50% formamide, 5 × SSPE, 1% SDS, 1 × Denhardt's reagent, 10% dextran sulphate, 10 µg ml⁻¹ salmon sperm DNA and random-primed (Boehringer-Mannheim) *mss4* cDNA 16 h at 42 °C¹⁸. The final wash was in 0.2 × SSPE and 0.1% SDS 1 h at 65 °C, and the blot was exposed 48 h with an intensifying screen.

a

```

1  GGTAGCCACTTTGGACTTATCTTCAAGAAATGTGACATTTGTGGTAATATTCTTTGCTTCC
61 CCATGGCCCTCACCCTGCCACAACAGTGTACGTGTTTGACACTCACTGCTTGTGCACCCA
121 TCACCTTCTGATGTATGGCAGTTTGGGTCAATAAGATATTAGTATTACTTTGACTTTGAC
181 TTGTGAGTCTAGGTTAAATCATTCGGAGGATTTTTATTCTCCGAGGTACCCCAACCG
241 AAATTTTATAGTTTATATTTATTTTGTGTTTAGCCATTAGGTTGTTTTATATAATTGAA
301 CTAGTAAATTAAGCTCCATAGGGCTTCTCCGTCTTATTGGGAGATTCCAGCCCTCTTAC
361 TGGAAAGGTCAATTTCACTGATTGAAAGTAAGAGACAGTTGAACCCCTCGTTTAGCCATTCA
421 TTCTAGTCCCTAATTAAGGAACAAGTGATTATGCTACCTTTGCACGGTCAGGATACCGCG
481 GCCGTTTAACTTTAGTCACTGGGCAGGCAATGCTCTAATACTTGTATGTAGAGGTATG
541 TTTTGGTAAACAGGCGGGGTTTCATGTTGCCGAGTTCCTTTTACTTTTAAATCTTTTCT
601 TAAAGCAGCCTGTGTTGGGCTAACGAGTTAGGGATAGGTAATCTATTGTTGGGTTAGT
661 ACCTATGATTCGATAAATGACAAATGGTTATCCGGGTTGTCATACACTTGTGCTAGGAGAG
721 GCAATGGAAACCTGCGAGCTGCAACAGAGCTAGTGTACGCCGAAGCCGGAACCGGAAG
1  MetGluProCysGluLeuGlnAsnGluLeuValSerAlaGluGlyArgAsnArgLys
781 GCGGTGCTGTGCCAGCGTTGTGGCTCTCGGGTGTGCAGCCAGGGACTGCTCTTTCTCT
20 AlaValLeuCysGlnArgCysGlySerGlySerGlySerGlySerGlySerGlySerGlySer
841 CGCCGGCAGCTTTTCTCTCCCTCCATGAGGAAGAAGCCGGATCTGGTGGACGGCAGCAAT
60 ArgArgGlnLeuPheLeuProSerMetArgLysLysProAspLeuValAspGlySerAsn
901 CCTGATGGTGTATGCTCTCGAGGAACACTGGCTGGTTAATGACATGTTTCATCTTTGAGAAC
80 ProAspGlyAspValLeuGluHisTrpLeuValAsnAspMetPheIlePheGluAsn
961 GTGGGCTTTACCAAGGACGTGGGCAACGTCAGTTTCTAGTCTGTGCAGACTGTGAGATT
80 ValGlyPheThrLysAspValGlyAsnValLysPheLeuValCysAlaAspCysGluIle
1021 GGACCCATTGGCTGGCATTGCCTAGATGACAAGAACAGCTTCTATGTGGCTTGGAGCGA
100 GlyProIleGlyTrpHisCysLeuAspAspLysAsnSerPheTyrValAlaLeuGluArg
1081 GTTTCACACGAGTAACAGGGGTTTGGGACTTGGCTCCACCTAGCTATAAAGCTCTGC
120 ValSerHisGluEnd
1141 CCACAAGAACCAGACTGATGTGTGAGCAGTGTGTCAGACTTCTTGGTACTAATATAAG
1201 CACCGGTGTTTGGCCTTGAACCTGCAGAGGGCCGGTCTGTTTCTCAGGCCCTCAGAGCA
1261 TGTCCTCTACTTAGTCTCTCTGTGCATGTGTCTAGAGCCCTTTCCCTGACTTGGAA
1321 CATGGAAGAACACGCGAGTGTCTACCACTGTAGTCTGGTCCGGCACTCTGCCTCCGCA
1381 CCTCGATGCCAAGCTCTCTTCTTCTCAGCAAGAACAATTCATCTTTAGCCATCACGAC
1441 GGGAACTAAGATGTCTAATATTGAACTCCACCTTCTCTCATCTAGAAGCCAAGTGTGA
1501 TAGGGTGCACATCTGTAATCTAGGTCTTGTAGTGGGTAGAGGCAGGAAGATCAGGAGTTC
1561 AAGGGCAGCCTTAGCTACATAGACTGGGCCAAGTGAGATGTTGTCTCAAAAACATAAAG
1621 CATACAAGTATTTGAGAGTAATCTCATGTATGTGACCTACTATCTACAGTAATTTATT
1681 CTTTTTAAAGTAAACATTTACAGATGTCTAGTATTTAAACCTCACTTTAGATACATTTT
1741 TAGTACTGTGTTTGAGAGCCTTTTTCCGGGTCAATAGTGGCATTTTATAATGAGGACGA
1801 ACACGACTCTCTTCCCTCAGATCTTGTACAGGTTTCTCTTTTGTGATCTCTTGTCCCT
1861 TGATGTCAACCCCTAAGGAAACTCAGCCCTAGCTGTGTGCAGACTCTCTTGTAAACCTG
1921 TTCCACGGACATTCTATTTGTATGGACAAGGTGTACCTTTTCTTATTATTAGCCAAC
1981 TAAATCTTCAATATGATTTCAGTTACTCTGTCTATTCTCTTCTGGAGTGACTCAGTAGC
2041 TTAAAGGACTCTGTGTAACACAGGTTCAAACTACTGTGTCATCATCCCTCTCTGAGGCG
2101 CTAAATTGAGATAAGTAGAGACTTCATTATTCTCATGCCAGAGCTTAGGAAGTGGGGGC
2161 AGATCAAGAGTTCAAGGTCAGCCTCAGTTACTATAATATCAAAATCAAGGCCAGCCTGTA
2221 CTATGGAGATCTTGTTCAAATGAACAAACACAAATAAACAAGTAGGACCTGCACAAAG
2281 TGACACGTAGTTAAATATGTTGTGCTCTTTTCAGGCGACCTGAGTTTCAGGACCTTGCAC
2341 CCTCATGGGAGGTTTATTCAGTTCAAGGAGATCTGCGCCTTCTGGTCCCTGCAGATGCC
2401 TTCAATGTGTGGCACCCATACAAACAGTAGGACACTCGCGTACACACACACACACGC
2461 ACGCACGCAGTGAATAAAAAATTTTACAAAAA

```

b**c** kb 1 2 3 4 5 6

9.5
 7.5
 4.4
 2.4
 1.4
 0.24

Growth at 37°

+
 +
 -
 -
 -
 -

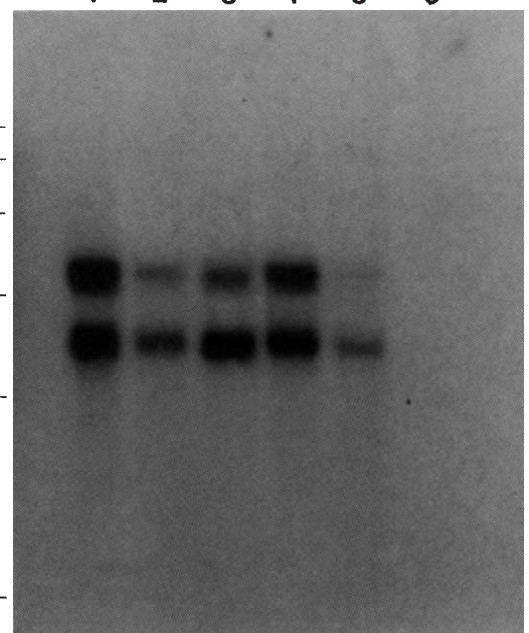


FIG. 3 Sequence comparison between the Mss4 and Dss4 proteins. Two regions of high homology, located at the carboxy termini of the two molecules, are boxed. An overall 27% identity and 51% similarity between the two proteins were calculated using the Bestfit computer program, which uses the local homology algorithm of Smith and Waterman²². Bars represent identical amino acids, and double dots are used when the amino-acid similarity threshold is greater than or equal to 0.50.

```

Mss4p 6 LQNELVSAEGRNRKAVLCQRCGRVLQPGTALFSRRQLFLPSMRKKPDLV 55
      : . . . | | : . : . : . : . : . : . : . : . : . : . : . :
Dss4p 5 MSKATCSFEGCHSAVITIN..DDNIINLPEQVHSEFKLL.....ENRTMR 47

Mss4p 56 DGSNPDGDVLEEHWLVNDMFIFENVGFTKDVGN..... 88
      | : . : . : | : : | : : | : | : . : . : . :
Dss4p 48 DATPSESNFL...VVPDVWDFDNGVSRIPSSILGDLSDKSDVFVEYG 93

Mss4p 89 .....VKFLVCADCEIGPIGWHLCLDDKNSFYVALERVSH 122
      : : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Dss4p 94 NSSWKIKKCLKYLICADCDKGPIGIICKVQDT..KNEERVLH 134
  
```

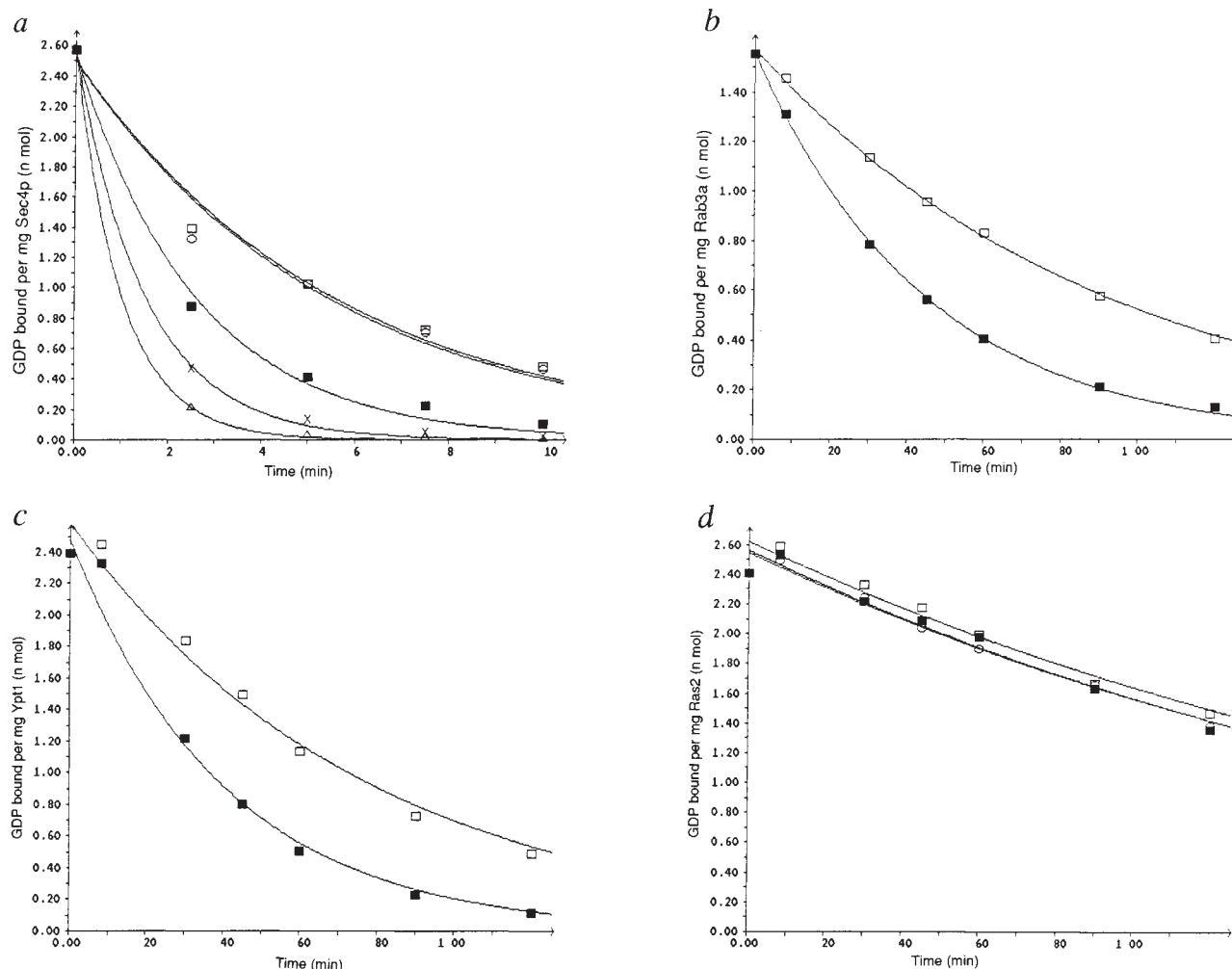


FIG. 4 The effect of various concentrations of Mss4-MBP on the rate of ³H-GDP release from recombinant GTP-binding proteins at a final concentration of 25 nM. *a*, Mss4-MBP, at 140 nM (Δ), stimulated the [³H]GDP off-rate of yeast Sec4 up to 5.5-fold over the intrinsic rate. The rate was unaffected by buffer ($\circ$) or MBP ($\square$). Mss4-MBP, at ~40 nM ($\blacksquare$), 2.2-fold stimulation; Mss4-MBP, at ~80 nM ($\times$), 3.7-fold increase. Data shown are the mean of duplicate determinations of two independent experiments. *b* and *c*, Mss4-MBP at 300 nM ($\blacksquare$) stimulated GDP release from Rab3a or YPT1 ~2-fold compared to release with MBP alone at 300 nM ($\square$). *d*, Mss4-MBP at 300 nM ($\blacksquare$) or 600 nM ($\circ$) had no effect on the GDP release rate from Ras2, which was identical to that with MBP alone ($\square$). For *b*-*d*, each data point was determined in triplicate and is representative of two identical experiments. METHODS. The complete ORF of the *mss4* cDNA was fused in-frame to the 3' end of the maltose-binding protein by cutting the 11A plasmid with *SacI* and *StuI* to subclone directionally into the pMal-p vector (NEB). A fusion protein of the expected size was obtained (56K; 14K for Mss4p and 42K for MBP). The Mss4-MBP fusion protein was purified by binding to an amylose resin column and elution of the protein with 10 mM maltose²³. The eluted protein was then dialysed against (in mM) either 10 Tris, pH 7.6, 100 NaCl or 50 HEPES, pH 8.0, 200 NaCl, 5 MgCl₂, 1 EDTA for *a*, and 50 HEPES, pH 8.0, 100 NaCl, 5 MgCl₂, 1 EDTA for *b*-*d*. Dialysed protein was concentrated to

1 mg ml⁻¹ using an Amicon Centriprep 10. Purity of ~90% was verified by 10% polyacrylamide-SDS electrophoresis followed by Coomassie brilliant blue staining. The filter assay was modified from ref. 11. Sec4 protein (100 nM) was incubated with 0.2 mM [8,5'-³H]GDP (Du Pont-New England Nuclear), at a specific activity of 20,000 c.p.m. pmol⁻¹ for 1 h at 30 °C. Purified Mss4-MBP, at various concentrations and supplemented with 1 mM cold GDP, was introduced into the Sec4-[³H]GDP mixture after the 60-min incubation. Final buffer was either buffer A (in mM, 40 HEPES, pH 8.0, 100 NaCl, 5 MgCl₂, 1 EDTA, 1 DTT, 0.5 GDP, 0.05% Lubrol 12A9, 0.05 mg ml⁻¹ BSA), or buffer A plus 162 μ M Tris. Results were similar in both buffers. Aliquots were taken at different time points and filtered over 0.45- μ m nitrocellulose filters (Millipore). Protein-associated radioactivity was assessed in a scintillation counter (Beckman LS 5000TD). Data were fitted to an exponential decay equation using nonlinear regression analysis²⁴ by the computer program Enzfitter. Yeast Sec4 was purified from an overproduced soluble pool as described¹¹. Filter assays for Rab3a, Ypt1 or Ras2 were done as for the Sec4 protein with the following modifications: the GTP-binding protein was incubated in 1 mM MgCl₂ with 2 mM EDTA for 30 min at 30 °C in the presence of ³H-GDP, then MgCl₂ was added to final concentration of 7 mM and incubated for 15 min at 30 °C.

member of the small GTP-binding protein superfamily. But when Mss4 was compared with a 17K yeast protein, Dss4 (ref. 7), a 27% identity and 51% similarity were found (Fig. 3). A point mutant of Dss4, Dss4-1, is a dominant suppressor of *sec4-8*. Both Dss4 and Dss4-1 proteins accelerate the GDP-off rate of Sec4 *in vitro*. The region containing the Dss4-1 mutation is the domain most highly conserved between the Mss4 and Dss4 proteins.

Because Mss4 and Dss4-1 proteins shared sequence homology and both genes showed a genetic interaction with *sec4-8*, we tested whether Mss4, like Dss4-1, could accelerate the intrinsic GDP off-rate of Sec4 *in vitro*. An Mss4/maltose-binding protein (MBP) fusion product (Mss4-MBP) was expressed in *Escherichia coli* and purified on an amylose resin column. The effect of this fusion protein on the rate of GDP dissociation from recombinant Sec4 protein purified from yeast was investigated using an *in vitro* filter-binding assay¹¹. Sec4 had an intrinsic off-rate of 0.21 min⁻¹ as measured by this assay. At a final concentration of 140 nM, Mss4-MBP accelerated the GDP off-rate of Sec4 5.5-fold (Fig. 4a). Identical results were obtained if Sec4 protein was purified from *E. coli* (data not shown). MBP had no effect on the intrinsic off-rate. These biochemical data suggest that the Mss4 protein rescues the *sec4-8* mutant phenotype by interacting directly with Sec4-8 protein to stimulate GDP-GTP exchange. To determine the specificity of Mss4, the releasing activity was tested on other GTP-binding proteins. At ~300 nM, Mss4-MBP accelerated the GDP off-rate twofold from both mammalian Rab3a (Fig. 4b) and yeast Ypt1 (Fig. 4c). In contrast, even at 600 nM, the Mss4-MBP had no effect on GDP release from yeast Ras2 (Fig. 4d).

Using a genetic approach in yeast, we have identified Mss4, a mammalian guanine nucleotide releasing protein which seems to act on members of the Sec4/Ypt1/*rab* branch of the *ras* superfamily. The Mss4 protein stimulates GDP release from both Ypt1 and Rab3a, but is less active on these proteins than on the Sec4 protein. These data indicate that in mammalian cells an endogenous substrate more closely related to Sec4 than to Rab3a might function as the preferred target for Mss4. The presence of the *mss4* mRNA in numerous tissues indicates that the Mss4 protein might play a general role in vesicular transport in all cells. Other guanine-nucleotide-releasing proteins that act on the Ras and Rho subgroups of the small GTP-binding protein superfamily have been reported^{12,13}. Preliminary characterization of a factor that stimulates GDP release from Rab3A has recently been reported¹⁴. The releasing proteins that have been cloned are larger than Mss4 or Dss4 (~60K), and the regions of these proteins thought to be involved in nucleotide release do not have significant homology with Mss4 or Dss4 (refs 12, 13, 15-17). Mss4 and Dss4 proteins therefore seem to represent a new family of guanine-nucleotide-releasing proteins. □

22. Smith, T. F. & Waterman, M. S. *Adv. appl. Math.* **2**, 482-489 (1981).
23. Kellerman, O. K. & Ferenci, T. *Meth. Enzym.* **90**, 459-463 (1982).
24. Marquart, D. W. *J. Soc. ind. appl. Math.* **11**, 431-441 (1963).

ACKNOWLEDGEMENTS. We thank T. Vernet for pVT102U, T. Sudhof for purified Rab3a, S. Powers for purified Ras2, M. Garrett for purified Ypt1, R. Bowser for advice on yeast spheroplast transformation and S. Amara's laboratory for help with cDNA library construction. This work was supported by grants from the NIH (P.N. and D.R.) and NIMH (P.D.C.).

Epithelial sodium channel related to proteins involved in neurodegeneration

Cecilia M. Canessa, Jean-Daniel Horisberger & Bernard C. Rossier*

Institut de Pharmacologie et de Toxicologie de l'Université,
Rue du Bugnon 27, CH-1005 Lausanne, Switzerland

THE epithelial amiloride-sensitive sodium channel constitutes the rate limiting step for sodium reabsorption by the epithelia lining the distal part of the kidney tubule, the urinary bladder and the distal colon. Reabsorption of sodium through this channel, which is regulated by hormones such as aldosterone and vasopressin, is one of the essential mechanisms involved in the regulation of sodium balance, blood volume and blood pressure¹⁻⁶. Here we isolate a DNA from epithelial cells of rat distal colon and identify it by functional expression of an amiloride-sensitive sodium current in *Xenopus* oocyte. The deduced polypeptide (698 amino acids) has at least two putative transmembrane segments. Expression of this protein in *Xenopus* oocytes reconstitutes the functional properties of the highly selective amiloride-sensitive, epithelial sodium channel. The gene encoding this rat sodium channel subunit shares significant sequence similarity with *mec-4* and *deg-1*, members of a family of *Caenorhabditis elegans* genes involved in sensory touch transduction and, when mutated, neuronal degeneration. We propose that the gene products of these three genes are members of a gene family coding for cation channels.

We used a functional expression assay in a *Xenopus* oocyte system⁷ to isolate and sequence a novel cDNA from a rat colon library. Figure 1 shows the nucleotide and deduced amino-acid sequence of the sodium channel clone, α rENaC. The 3,161-nucleotide (nt) clone contains a 2,094-nt open reading frame that codes for a 698-amino-acid protein with a relative molecular mass (M_r) of 79K. The rat clone hybridized to a 3.5 kilobase (kb) transcript on northern blots containing rat poly(A)⁺ messenger RNA isolated from kidney cortex and medulla, the distal part of the colon and the stomach (Fig. 2). Transcripts were not detected in brain, liver, jejunum and proximal colon. The tissue distribution of this transcript agrees with that of a highly selective sodium channel in mammals¹. The weak gastric signal could correspond to the amiloride-sensitive sodium transport observed in gastric mucosa of newborn pigs and rabbits⁸.

Oocytes injected with complementary RNA (cRNA) from the rat clone expressed an amiloride-sensitive current (Fig. 3a, c, upper panels) not present in the water-injected controls (Fig. 3b, d). The ion selectivity of the channel was assessed by replacing Na⁺ with K⁺ in the oocyte bathing solution. No amiloride-sensitive current could be detected in K⁺ buffer. The current-voltage relationship of the amiloride-sensitive current (Fig. 3e) indicated inward rectification as expected from the inwardly directed sodium gradient. The amiloride sensitivity was measured at -100 mV in Na⁺ buffer (Fig. 3f). The estimated K_i was 0.1 μ M which agrees with previous values using similar experimental conditions¹⁻³.

The amiloride-sensitive sodium channel has been isolated

Received 19 October; accepted 27 November 1992.

1. Hall, A. *Science* **249**, 635-640 (1990).
2. Balch, W. *Trends biochem. Sci.* **15**, 473-477 (1990).
3. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **349**, 117-127 (1991).
4. Pfeffer, S. *Trends Cell Biol.* **2**, 41-45 (1992).
5. Bollag, G. & McCormick, F. *Rev. Cell. Biol.* **7**, 601-632 (1991).
6. Salminen, A. & Novick, P. *J. Cell Biol.* **49**, 527-538 (1987).
7. Moya, M., Roberts, D. & Novick, P. *Nature* **361**, 460-463 (1993).
8. Vernet, T., Dignard, D. & Thomas, D. *Gene* **52**, 225-233 (1987).
9. Novick, P. J., Field, C. & Schekman, R. *Cell* **21**, 205-215 (1980).
10. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
11. Kabaceni, A. K., Goud, B., Northrup, J. & Novick, P. *J. J. Biol. Chem.* **265**, 9366-9372 (1990).
12. Jones, S., Vignais, M. & Broach, J. R. *Molec. cell. Biol.* **11**, 2641-2646 (1991).
13. Kaibuchi, K. *et al. Molec. cell. Biol.* **11**, 2873-2880 (1991).
14. Burnstein, E. S. & Macara, I. G. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1154-1158 (1992).
15. Chant, J., Corrado, K., Pringle, J. R. & Herskowitz, I. *Cell* **65**, 1213-1224 (1991).
16. Powers, S., Gonzales, E., Christensen, T., Cubert, J. & Broek, D. *Cell* **65**, 1225-1231 (1991).
17. Simon, M. A., Bowtell, D. D. L., Dodson, G. S., Lavery, T. R. & Rubin, G. M. *Cell* **67**, 701-716 (1991).
18. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
19. Burgers, P. M. J. & Percival, K. J. *Analyt. Biochem.* **163**, 391-397 (1987).
20. Ito, H., Fukada, Y., Murata, K. & Kimura, A. *J. Bact.* **153**, 163-168 (1983).
21. Davis, L. G., Dillner, M. D. & Battey, J. F. *Basic Methods in Molecular Biology* (Elsevier, New York, 1986).

* To whom correspondence should be addressed.

-82 CCGGGTCGACCCACGCGTCGCGAGGGCAGCCTGGGATCGGGCAGGTCGCCGACAGCCCCATTCTGCCTTACGCTAATG -1

ATGCTGGACACACAGAGCCCTGAGCTCAACATTGACCTAGACCTTACGCTTCAACTCGCTAAGGGGTCATGAAGGGCAACCAATCAAGGAGCAAGACCTTGTCTCTCAG 120
M L D H T R A P E L N I D L D L H A S N S P K G S M K G N Q F K E Q D P C P P Q 40

CCCATGCAAGGACTGGGAAGGGGACAAACGTGAAGAGCAGGGCCTGGGCCCGGAACCTCAGCACCCCGGACGCCAGGAGGAGGAGGACTGATTGAATTCACCGCTCCTAC 240
P M Q Q L G K G D K R E E Q G L G P E P S A P R Q P T E E E A L I E F H R S Y 80

CGGGAGCTCTTCCAGTCTTCTGCAACAACACCACCTCCAGGGGCCATCCGCTGGTGTGTCTCAACACAACCGCATGAAGACGGCTTCTGGCGGTGCTGTGGCTGTGCACCTTC 360
R E L F Q F F C N N T T I H G A I R L V C S K H N R M K T A F W A V L W L C T E 120

GGCATGTACTGGCAGTTGCGCTTGTCTGAGGAGTACCTCAGCTACCCAGTGAAGCTCAACATCAACCTCAATTGAGACAAGCTGGTCTTCCCTGCCGTCAGTGTCTGCACCTT 480
G M M Y W Q F A L L F E E Y L S Y P V S L N I N L N S D K L V F P A V T V C T L 160

AATCCTTACAGATACACTGAAATTAAGAGGAGCTGGAAGAGCTGGAC: 3CATCACGGAGCAGACGCTTTTGTACTGTACAAATACAACCTTCTTACCTCGCCAGGCTGGGGCCGCA 600
N P Y R Y T E I K E E L E E L D K I T E Q T L F D L Y K Y N S S Y P R Q A G A R 200

CGCCGAGCTCCCGGACCTCCTGGTGTCTTCCCGCACCCCTCGAGCGCTCGGCACTCCACCTCCGCTTACTCCGGCCGACGGCGCAGCGGTCTTCAGCGTACGGCAGCAACA 720
R R S S R D L L G A F P H P L Q R L R T P P P Y S A G R T G A A G L P A Y A T T 240

ATCCCCCAAGTGACCGGAAGGACTGGAAGATCGGCTTCAACTGTGCAACAGCAACAATCAGACTGTTTCTACAGACATACCTCTGGGTGGATGAGTGAGGGAGTGGTACCGC 840
I P Q V D R K D W K I G F Q L C N Q N K S D C F Y Q T Y S S G V D A V R E W Y R 280

TTCCATTACATCAACATTCTGTCCAGACTGTGCGACACCTCGCCGCTCTAGAGGAAGAAGCCCTGGGCACTTCACTTTCACCTGTGCTTCAACAGGCCCCCTGCAACAGGCGAAT 960
F H Y I N I L S R L S D T S P A L E E E A L G N F I F T C R F N Q A P C N Q A N 320

TATTCAGTTCACACCCCATGTACGGGAAGTCTGCTACACTTCAATGACAAGAACTCAATCTCTGGATGTCTCCATGCCTGGAGTCAACAATGGTTTGTCTCCGACACTGCGC 1080
Y S K F H N F M Y G N C Y T F N D K N N S N L W M S S M P G V N N G L S L T L R 360

ACAGAGCAGAGTACTTCACTCCCTGCTGTCACAGTACGCGGCGCAGGGTGTGGTGTGTCAGGATGAGGCTGCTTATGATGATGGTGGCTTCAACTTGAGGCTGGCGT 1200
T E Q N D F I P L L S T V T G A R G V M V H G T G D E P A F M D D G G F N L R P G V 400

GAGACCTCCATCAGTATGAGAAGGAAGCCCTGGACAGCCTCGGAGGAAATTACGGCGCACTGTACTGAGAATGGTAGCGATGTCCCGTCAAGAACCTTACCTTCAAGTATACACAG 1320
E T S I S M R K E A L D S L G G N Y G D C T E N G S D V P V K N L Y P S K Y T Q 440

CAGGTGTGCTTACCTCTGCTTCCAGGAGAACATGATCAAGAAGTGTGGCTGTGCTACATCTTACCTTAAGCCCAAGGAGTTGAGTTCTGTGACTACCGAAAGCAGAGCTCCTGG 1440
Q V C I H S C F Q E N M I K K C G C A Y I F Y P K P K G V E F C D Y R K Q S S W 480

GGCTATTGCTATTATAAATGCAGGGCGCTTCTCTTGGACAGCCTGGGCTGTTTCTCAAGTGTGGAAGCCTTGTAGTGTGATCAACTACAACCTCTCTGCCGGTACTACGGTGG 1560
G Y C Y Y K L Q G A F S L D S L G C F S K C R K P C S V I N Y K L S A G Y S R W 520

CCATCTGTGAAGTCCAGGATGGATCTTGGAGTGTCTTGCAGAACATTAACATAACAAAGAAACGAGTTCGAAAGCTCAACATCTTCTCAAGGAGCTGAAGTAT 1680
P S V K S Q D W I F E M L S L Q I N Y T I N N K R N G V A K L N I F K E L N Y 560

AAAATAATTCGGAGTCTCTTCTGTGACGATGGTCAAGCTCTGTCTCA:CTGGGACGCGAGTGGAGCTGTGGTTGGCTGTCTGCTGTGGTGGAGATGGCGGAGCTATC 1800
K T N S E S P S V T M V S L L S N L G S O W S L W F G S S V L S V V E M A D V I 600

TTGACCTCTGTGCTATCAGCTTCTCATGCTGTACGCGGTTCCGGAGCGGTACTGTCTCCAGGACGAGGGGCGCAGGAGTGGCTCCAGCTCCAGCTTCTCTCTT 1920
D L L V I T L L M L L R R F R S R Y W S P G R G A R G A R E V A S T P A S S F 640

CGTCCGTTTCTGTCTCACCTACATCCACACCTTCTTTGCCAGCAGGGGATGACCCCTCCCTGGCCCTGACAGCCCTCCACCTGCCTATGCTACTAGGCCCACTGCC 2040
P S R F C P H P T S P P P S L P Q Q G M T P P L A L T A P P P A Y A T L G P S A 680

CCTCCTGAGTCTGCGGCGCTGACTGTTCTGCTGTGCTGCGGCGCTCTGAGAGGAGGAAGATCCTCTCACCCAGCCCTGAGCTCCCTTGTAAACTGATGGATATCTCACTT 2160
P P L D S A A P D C S A C A L A A L STOP 698

CAGCACATCTTCCCGAGCGCCCGCGGAGCCCTCTGTGGTGTGCTGCTGAGGGAACAGGCTGAGTAAGGGGCCAGGAAGCCGTCAGAGGACAGGGGCTAACGATCTGCTCAGAGC 2280
TGCCCTGCCCTGCTTCTGAACACTGCCTTCCACAAGCACAGACAAGTCCCTCTTCCCTCGGATCAGCCAGCCAGACTTGGAGCTCTGACAAGAACTTCTCTGGGAGAGGACCAAG 2400

AACAATAGAAGCCATCGCACTCTACCCCGGTGACGAGGCTGGCTGGCTCACTGCCTTACGGGACACTGATGTCTGCTCTCTCTGAACTTGGTGGGAACCCACCCAA 2520
AAGCCCCCTGTTAGTCTTGGCAATTCCTCTCCCTCACTCTCAGGTGGGACTAGAGTAAGACAGACAGGGTAGGCTGGTGTCTCTCTCATTCTCAAGACTCTCTCATTGA 2640

TACACCATACCCAGTGCCTCTGTGGTGTGTCTCTCTTCTTCTGTTTCTCAGCTGGAAGCTTCTTCTACTGTCCCTCAGAACACAGCTCTATGCATCCCTTAGAAACCTGCCAAA 2760
GGCACCATTCTTTGTGAAGTCACTTACCTGTTATGTCTCCCCAGAATGTGCGCTCTCCCTCACCCCAAGTGTACTGTATTATATGTTACATCTCTTTGTGTCTGCCTCCC 2880

TGGGTAACTGAGCTCCTTGTGTGGTCAAGGATGAGATTGCTCTGTTTGTATCTTCCGGTCTAGCCAGTCTCCCACTTGGGACGGGTAGGAGGTACTCAATAAAGGCTTGTTC 3000
TCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3079

FIG. 1 Nucleotide sequence of cDNA and deduced amino acids of α ENaC. Nucleotides are numbered, starting at the first residue of the putative initiator ATG, according to Kozak's consensus sequence (no stop codons are found in the upstream sequence). Putative transmembrane spanning domains are underlined. Potential sites for *N*-linked glycosylation are indicated by asterisks, and the consensus sites for the cAMP-dependent phosphorylation by circles. Potential nucleotide-binding site is underlined by a dashed line. The amino terminus (Met 1–Ala 110) contains two potential sites for glycosylation and an important cluster of negatively charged amino acids. It includes also a consensus sequence involved in nucleotide binding (Gly-X-Gly-X-Gly)²². A first long hydrophobic domain, H1, could span the membrane at least once. A large, mostly hydrophilic domain (Asn 161–Ser 568) has six potential *N*-glycosylation sites. The 3 potential cyclic nucleotide-dependent protein kinase phosphorylation sites (Ser 203, Ser 204, Ser 478), may play a role in the vasopressin-mediated regulation of

the sodium channel⁵. A second large hydrophobic domain H2 (Val 569–Leu 612) could span the membrane at least once. Finally, a cluster of positively charged residues (Arg 613, 614, 616) is present in the mostly hydrophilic carboxy terminus.

METHODS. Poly(A)⁺ RNA was extracted from the distal colon mucosa of rats kept on a low salt diet for one week in order to stimulate sodium reabsorption. Poly(A)⁺ RNA was size-fractionated in a sucrose gradient⁷. The fraction that induced maximal amiloride-sensitive Na⁺ current in injected *Xenopus* oocytes (corresponding to 18S) was used to construct an oligo(dT)-primed directional library in pSPORT (BRL). 60,000 independent clones were obtained and initially divided into 10 pools. RNA was transcribed from each pool of cDNA and screened for its ability to induce amiloride-sensitive Na⁺ current when injected into *Xenopus* oocytes. After several rounds of division, a single functional clone, α ENaC, was isolated. DNA sequence of both strands was determined by the chain-termination reaction method using Sequenase.

from bovine renal papilla and from an amphibian kidney cell line (A6). The native protein has an apparent M_r of ~ 730 K. Upon disulphide bond reduction and denaturation of the complex, five or six polypeptides with an apparent M_r of 320K, 150K, 95K, 70K, 55K and 41K were observed^{9,10}. We propose that the clone encodes one of the subunits of the epithelial

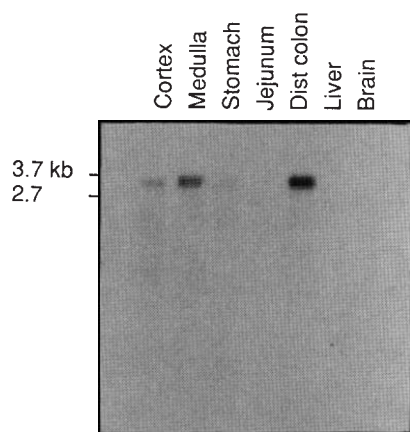


FIG. 2 Northern blot analysis of α ENaC expression in various rat tissues. Molecular mass markers are shown on the left.

METHODS. Poly(A)⁺ mRNA was electrophoresed on a 0.8% denaturing glyoxal agarose gel and transferred to a nylon membrane. The membrane was hybridized with random-primed ³²P-labelled dATP full-length α ENaC cDNA. 3 μ g mRNA was loaded for all tissues except colon, for which 1 μ g was loaded.

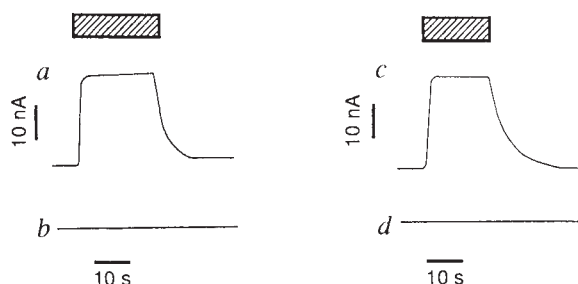


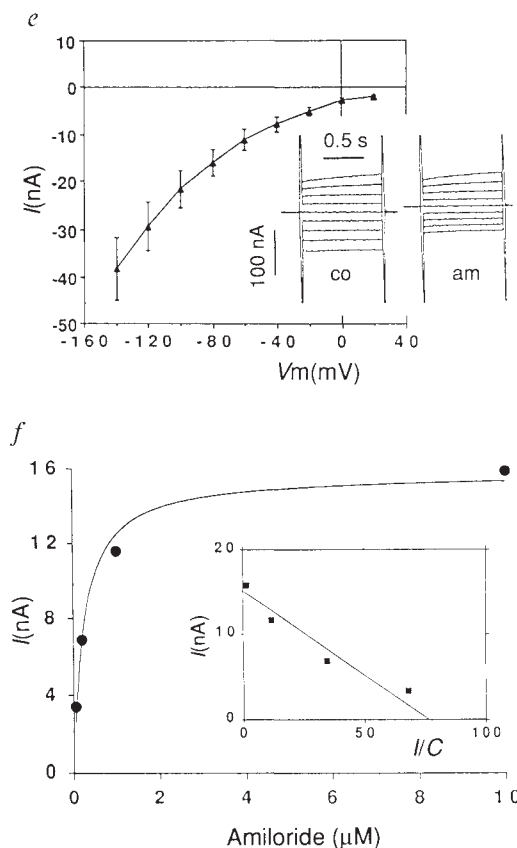
FIG. 3 Amiloride-sensitive current in *Xenopus* oocytes. *a* and *b*, current recordings showing the effect of 2 μ M amiloride (hatched bar) on the whole-cell current at -100 mV holding potential in *a*, an oocyte injected with 10 ng α ENaC cRNA, and *b*, an oocyte injected with a similar volume (50 nl) of water. *c* and *d*, Ionic selectivity of the amiloride-sensitive current. Current recordings showing the effect of 2 μ M amiloride on the whole-cell current in an oocyte injected with α ENaC cRNA, first (*c*) in a Na^+ -containing buffer and second (*d*) in a Na^+ -free buffer (K^+ instead). *e*, Current-voltage relation of the amiloride-sensitive current. The amiloride-sensitive current is the difference between the current before and after addition of 10 μ M amiloride to the bathing solution. Values are the mean ($\pm$ s.e.) of 6 measurements. Inset: original current recordings obtained without (co) and with (am) 10 μ M amiloride in an oocyte injected with α ENaC cRNA. *f*, Amiloride dose-response curve at a membrane potential of -100 mV in an oocyte injected with α ENaC cRNA. Inset: Eadie-Hofstee plot used to estimate the inhibition constant (K_i). Amiloride was added at increasing concentration from 0.0 to 0.05, 0.2, 1.0 and 10 μ M. The mean ($\pm$ s.e.) K_i was 0.10 ± 0.03 μ M ($n=4$) at -100 mV.

METHODS. *Xenopus* oocytes were injected with 10 ng α ENaC cRNA in 50 nl. Electrophysiological measurements were made 2 days after cRNA injection. Amiloride-sensitive current was measured using the two-electrode voltage-clamp technique⁷. The current signal was low-pass filtered at 25 Hz. The current-voltage relation (*e*) was determined by setting the command potential for 1 s (at 4-s intervals) from a -60 mV holding potential to values ranging from $+20$ to -140 mV. The bathing solution contained the K^+ channel blockers barium and TEA, and had a low chloride concentration to

sodium channel complex. Depending on the extent of glycosylation, the 78K translation product could give rise to the observed 95K or 150K species. As the *Xenopus* oocyte does not express endogenous sodium channel activity⁷ (Fig. 3), we propose that the cDNA encodes a pore-forming subunit of the channel protein that we shall name α ENaC (for α -subunit of the rat epithelial Na^+ channel).

Although a homomeric complex was sufficient for detection of the sodium channel activity in the oocyte assay, the subunit composition of the native channel is unknown. The Na^+ current observed after α ENaC cRNA injection in *Xenopus* oocytes was smaller than that observed after injection of size-fractionated poly(A)⁺ mRNA isolated from the tissue of origin. The reasons for this small current are not clear. We suspect that other subunits may play a role in modulating the expression of the channel in the plasma membrane or the activity of the pore-forming subunit, as shown for the calcium channel¹¹. Recently the primary structure of an apical protein (Apx) (1,420 amino acids) from *Xenopus laevis* that participates in amiloride-sensitive sodium channel activity was reported¹². Complementary RNA of the Apx cDNA injected into the oocyte, however, failed to reconstitute amiloride-sensitive sodium current. In *Xenopus* kidney cells, the activity of a Na^+ channel is under the control of the α_{i-3} G protein¹⁰. The Apx protein and the α_{i-3} G-protein subunit are likely candidates for this modulatory role.

Comparison of the sodium channel subunit sequence against those of the EMBL/Gen Bank showed highly significant sequence similarities with *C. elegans*, *deg-1* and *mec-4*, two genes which arose from a genetic screen for touch-insensitive mutants¹³⁻¹⁵. Figure 4 compares α ENaC with these *C. elegans* genes in a stretch of 270 amino acids near the carboxy terminus



reduce the membrane conductance. The bathing solution contained (in mM) Na^+ (87), Ca^{2+} (0.4), Mg^{2+} (0.82), Ba^{2+} (5.0), TEA⁺ (10.0), gluconate (80), Cl^- (22.5), HCO_3^- (2.4), MOPS buffer (10.0), pH 7.4.

of the protein, in which the similarities are most obvious. This region corresponds to the second hydrophilic domain and to the second hydrophobic domain (H2) of the sodium channel subunit. As the amino-terminal regions of *deg1* and *mec-4* have not yet been cloned, the overall similarity between the three proteins cannot be evaluated.

The *mec-4* and *deg-1* genes have been named 'degenerins' because dominant mutations of *mec-4* alanine 442 (Ala to Val) located within the second large hydrophobic domain (H2) are sufficient to induce selective degeneration of a few sensory neurons of the nematode. The phenotype of this mutant is striking: the cell bodies of degenerating neurons swell, reaching several times their normal cell diameter, before they disappear, presumably by cell lysis¹⁴. By contrast, conservative substitutions in which the amino-acid side chains are smaller or similar in size to alanine (Gly, Ser, Cys) did not produce the abnormal phenotype. The sodium channel α ENaC has a serine residue (Ser 578) in the corresponding position which should not modify significantly the secondary structure of the protein in this domain. It has been suggested that *mec-4* and *deg-1* genes code for "membrane proteins, that could regulate membrane transport, perhaps as components of ion channels or transport systems"^{14,15}. The function of *deg-1* and *mec-4* proteins could not be assayed as both clones were not full-length. The high

degree of conservation between an essential component of the mammalian epithelial channel and the nematode *deg-1* and *mec-4* proteins strongly suggests similar functions. If *mec-4* and *deg-1* are also components of a cation channel complex, the dominant mutation could lead to an abnormal permeability of the channel protein. The enhanced entry of NaCl and water into the cell could explain the cell swelling, vacuolization and finally cell death.

Considering the role in mechanosensation attributed to *deg-1* and *mec-4* gene products and the disturbance of cell volume control when they are mutated, we propose that these proteins and the rat epithelial sodium channel belong to a gene family encoding cation channels involved in the control of the cellular and extracellular volume. This family could also include mechanosensitive ion channels found in bacterial, yeast and animal cells¹⁶⁻²¹. □

Received 21 September; accepted 22 December 1992.

- Palmer, L. G. *Rev. Physiol.* **54**, 51-66 (1992).
- Garty, H. & Benos, D. J. *Physiol. Rev.* **68**, 309-372 (1988).
- Eaton, D. C. & Hamilton, K. L. in *Ion Channels* (ed. Narahashi, T.) 251-282 (Plenum, New York, 1988).
- Rossier, B. C. & Palmer, L. G. in *The Kidney, Physiology and Pathophysiology* (eds Seldin, D. W. & Giebisch, G.) Vol. 2, 1373-1409 (Raven, New York, 1992).
- Schafer, J. A. & Hawk, C. T. *Kidney Int.* **41**, 255-268 (1992).
- Horisberger, J.-D. & Rossier, B. C. *Hypertension* **19**, 221-227 (1992).
- Palmer, L. G., Corthésy-Theulaz, I., Gaeggeler, H.-P., Kraehenbühl, J.-P. & Rossier, B. J. *gen. Physiol.* **96**, 23-46 (1990).
- Machen, T. E., Silen, W. & Forte, J. G. *Am. J. Physiol.* **234**, E228-E235 (1978).
- Benos, D. J., Saccomani, G. & Sariban-Sohrab, S. J. *biol. Chem.* **262**, 10613-10618 (1987).
- Ausiello, D. A., Stow, J. L., Cantileo, H. F., de Almeida, J. B. & Benos, D. J. *J. biol. Chem.* **267**, 4759-4765 (1992).
- Singer, D. et al. *Science* **253**, 1553-1557 (1991).
- Staub, O. et al. *J. Cell Biol.* (in the press).
- Driscoll, M. & Chalfie, M. *Trends. Neurosci.* **15**, 15-19 (1992).
- Driscoll, M. & Chalfie, M. *Nature* **349**, 588-593 (1991).
- Chalfie, M. & Wolinsky, E. *Nature* **345**, 410-416 (1990).
- Martinac, B., Buechner, M., Delcourt, A. H., Adler, J. & Kung, C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2297-2301 (1987).
- Gustin, M. C., Zhou, X. L., Martinac, B. & Kung, C. *Science* **242**, 762-765 (1988).
- Guharay, F. & Sachs, F. J. *Physiol.* **352**, 685-701 (1984).
- Sigurdson, W. J. & Morris, C. E. J. *Neurosci.* **9**, 2801-2808 (1989).
- Hudspeth, A. J. *Hear. Res.* **22**, 21-27 (1986).
- Ohmori, H. J. *Physiol.* **350**, 561-581 (1984).
- Paulmichl, M. et al. *Nature* **356**, 238-241 (1992).

ACKNOWLEDGEMENTS. We thank J.-P. Kraehenbühl, W. Wahli, J.-M. Mathieu and R. Felder for their comments, M. Peitsch for help with the BLAST program, and N. Skarda-Coderey for secretarial assistance. This study was supported by grants from the Swiss National Foundation for Scientific Research.

A novel intron site in the triosephosphate isomerase gene from the mosquito *Culex tarsalis*

Claus Tittiger, Steve Whyard & Virginia K. Walker

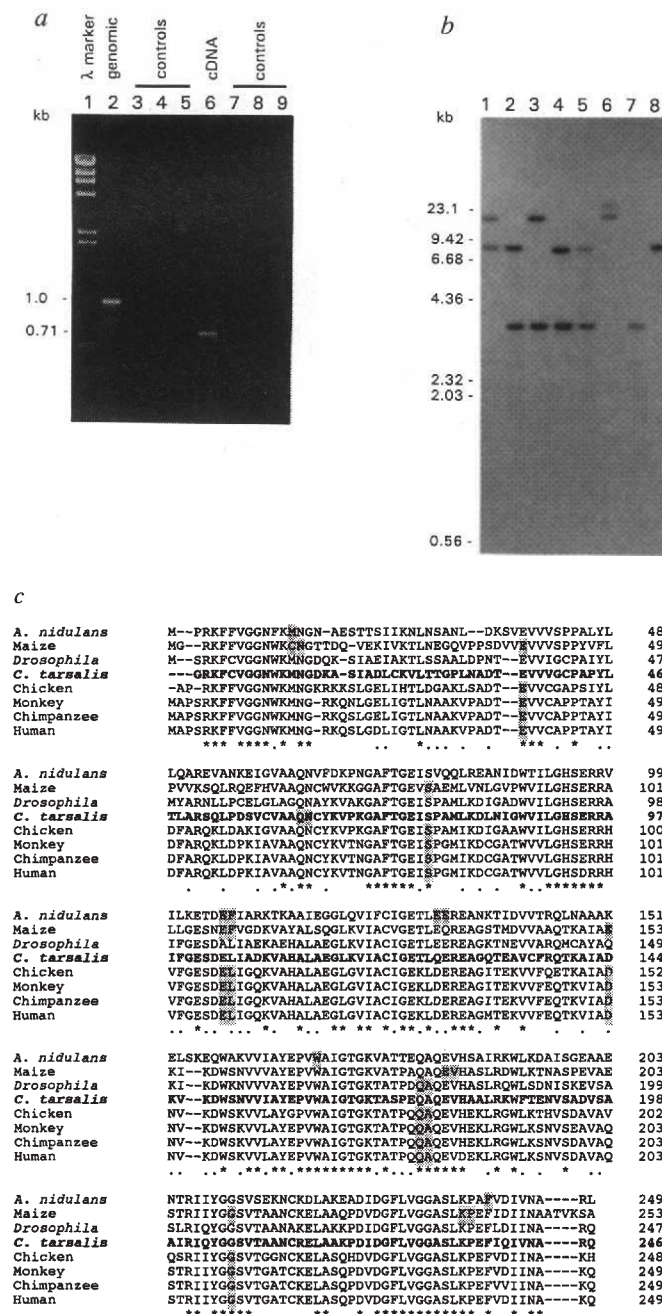
Department of Biology, Queen's University, Kingston, Ontario, Canada K7L 3N6

THE origin and function of introns in eukaryotic genes has provoked considerable debate since their discovery in 1977. Central to this issue are studies on the highly conserved enzyme, triosephosphate isomerase (TPI, EC 5.3.1.1). The 'introns early' argument suggests that introns are as old as the genes themselves and that the apparent correlation of many of the intron sites in plant, animal and fungal TPI genes with the boundaries of modules¹ is evidence of the assembly of ancient proteins by exon shuffling²⁻⁴. In contrast, the 'introns late' view holds that ancient genomes contained few if any introns; introns were inserted into pre-existing genes during the last billion years^{5,6}. We have found that the TPI gene from the mosquito, *Culex tarsalis*, contains an intron in a unique position that was predicted by W. Gilbert² and the exon shuffling hypothesis.

TPI was purified to homogeneity from fourth instar larval homogenates by successive steps of ion exchange, ultrafiltration,

Mec-4	LRMLVYVNADSYMPT-TEATGVRI fIHDKEDFFFPDFTGY	
α ENaC	LSLTLRTEQNDFIPLLSTVTGARVMVHGQDEPAFMDGGF	394
Deg-1	IRVLLFVNTSDYMST-SESSGVRLAIHPPTTEYFPDFTGY	
Mec-4	SAPTGYVSSFGLRLRKRSLPAPYGCDCVPDGKTS--DYIY	
α ENaC	NLRPGVETSIISMRKEALDSLGGNYGDCETENGSDVPVKNL	434
Deg-1	SAPVGFASSFGIKKKVMQRLPAPYGCCEVETKKVVDNRNIY	
Mec-4	SNYEYSVEGCRYSCFQQLVLKECRGDPFRFPVENARHCD	
α ENaC	PS-KYTQQVCIHSCFQENMIKKCGCAYIFYPKPKGVFECD	473
Deg-1	AGYDYHPEGCRSCFQNLIDDCSGDPFRFPVEGYRHCS	
Mec-4	AADPIARKCLDARMNDLGG-LH--GS FRCRC--QQPCRQ	
α ENaC	YRKQSSWGVCYCYKLQGAFS-LDSLGC-FS-KC--RKPCSV	508
Deg-1	AFNATART-CLEKNIGSVGDFHHITQKMD-KCVCKQSCCE	
Mec-4	SIYSVTYSPAKWPSSLQIQLGSCNGTAVECNKHYKENG-	
α ENaC	INYKLSAGYSRWPSVKSDWIFEMLSLQNNYTINNKRNGV	548
Deg-1	IIHEVTFSCSKWPSGATD--LGDCDGMTSECEQYRLNA	
Mec-4	AMVEVFYEQLNFEMLTSEAYGEV LLADFGGOLGLWCGL	
α ENaC	AKLNIFFKELNYKTNSPSTVMTVSLLSNLGSOWSLWFGS	588
Deg-1	AMIEVFYEQLNYELLQSEAYGLVNLTADEFGHGLGLWLG	
Mec-4	SFLTCCFEVFLFLETAYMSAENYSLYKKKKAEEKKAKIAS	
α ENaC	SVLSVVEADVIFDLVITLMLLRFRSRYWSPGRGARG	628
Deg-1	SVITVMEVCVLLVDMISLFEKSRHEEKLLRQSTKRKDVPE	

FIG. 4 Alignment of α ENaC with *mec-4* and *deg-1* protein sequences. Pairwise comparisons (rat protein with *mec-4* or *deg-1*) are shown. Amino acids of *deg-1* and *mec-4* identical to α ENaC are indicated by double dots. Single dots indicate conservative changes. Numbering refers to the rat sequence. The H2 hydrophobic domain is underlined. Sequences were searched at the National Center for Biotechnology Information (NCBI) using the BLAST network service.



According to the 'ancient introns' form of the exon shuffling hypothesis, primitive exons coded for modules which are defined by M. Gö¹ as peptides that assume a compact structure usually less than 2.8 nm across within a globular domain. These were shuffled by crossover at flanking intron sites or by retrotransposition of transcriptional products to produce novel proteins⁷⁻⁹. Given the remarkable correlation of known intron sites with modular boundaries in identified TPI DNAs, it was hypothesized that an ancestral gene was interrupted by 11 or 12 introns^{2,10}, 10 of which have been identified. To explain the existence of exon III, which appeared to encode two modules, it was predicted by W. Gilbert and his colleagues that a missing intron would eventually be discovered². We have now found

FIG. 1. Isolation and analysis of *C. tarsalis* TPI sequences. *a*, PCR amplification of genomic and cDNA. The 1,005 bp band in the genomic lane indicated the presence of at least one intron in the sequence. *b*, Southern blot analysis of *Eco*RI-digested genomic DNA isolated from individual adult mosquitoes hybridized with TPI cDNA. There are four morphs in our population of mosquitoes¹⁶, lanes 1–6 are heterozygotes, 7 and 8 are homozygotes. *c*, Amino-acid sequences (single-letter code) and intron positions of all TPIs for which genomic sequences containing introns are available. The deduced *C. tarsalis* sequence was aligned with TPI sequences from *Aspergillus nidulans*¹⁷, maize³, *D. melanogaster*^{1,2}, chicken⁴, rhesus monkey¹⁸, chimpanzee¹⁹ and human²⁰ using the CLUSTAL program provided by PCGene version 6.0 (IntelliGenetics). Intron sites are indicated by shading over an amino acid where the site interrupts the codon, or over two amino acids where the site falls between the codons. Conserved amino acids are indicated with asterisks beneath the sequences, similar amino acids are indicated by dots. Gaps inserted to optimize alignments are indicated by dashes. The amino-acid sequence is available from Genbank under accession number L07390. **METHODS.** *a* Genomic DNA was isolated²¹ from 30 decapitated adult *C. tarsalis*. Total RNA was isolated²² from fourth instar larvae and first-strand cDNA was prepared from oligo-dT primed RNA with SuperScript MMLV RNase H⁻ reverse transcriptase (Bethesda Research Laboratories) following the manufacturer's protocol. Genomic DNA (0.4 µg) and 0.25 µg cDNA were used as templates in separate reactions. Amplifications were done using a 64-fold degenerate sense oligonucleotide, TPI-2, encoding residues 9 to 16 of the N-terminal sequence (TGGAAUATGAAYGGNGAYAAUGC), and a specific antisense oligonucleotide, TPIREV, corresponding to amino acids 240–246 (GTCTGGCGTTGACAATCTGG). Forty cycles of amplification (1 min at each 94 °C, 60 °C, and 72 °C; followed by a 72 °C, 15 min extension) were done using Taq polymerase and standard buffer provided by the supplier (Promega). Abbreviations in the nucleotide sequence are N=A, C, G, T; U=A, G, Y=C. *b*, DNA was digested with *Eco*RI, electrophoresed, transferred and alkali-fixed to Hybond N⁺ nylon membrane (Amersham), and subsequently hybridized to the ³²P-labelled cDNA fragment in 50% formamide and washed under high stringency²³. *c*, A cDNA library was constructed from poly(A)⁺ RNA isolated from fourth instar larvae in λ ZAP II following the manufacturer's specifications (Stratagene). Template DNA for anchored PCR was isolated from 10⁶ phage and amplified using the T7 primer from pBluescript SK- (Stratagene) and TPI-2 as in *b*, but with an annealing temperature of 63 °C. Sequence information from a clone isolated by screening 10⁵ phage with the PCR product by standard methods²³ was used to design TPIREV for use in PCR amplification of genomic sequences. The 1,005 bp PCR fragment was isolated and ligated into the *Sma*I site of pBluescript SK + (Stratagene). The sequence of the 5' end of the gene missing from the cDNA was provided using a fragment isolated from a genomic library. All sequencing reactions were done at least twice on a 373A automatic DNA sequencer (Applied Biosystems) using dye terminators.

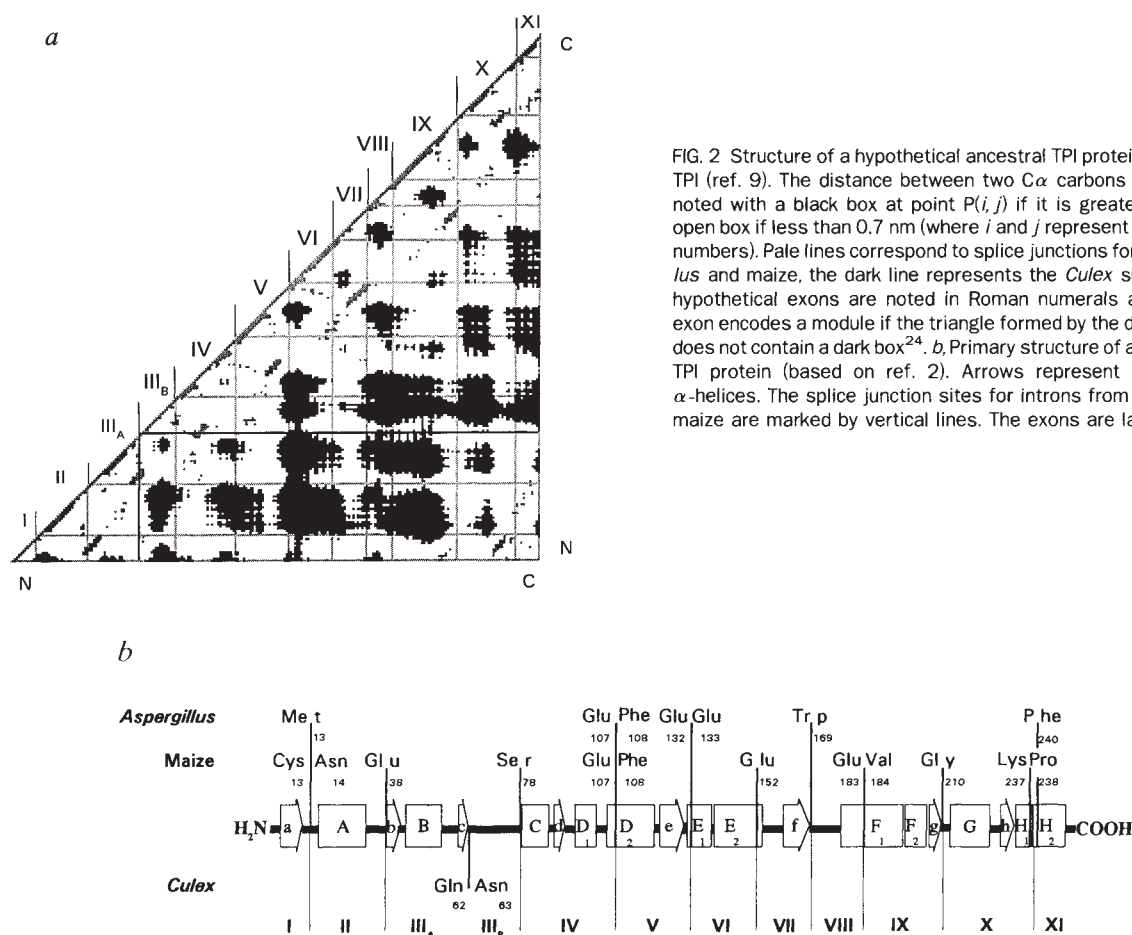


FIG. 2 Structure of a hypothetical ancestral TPI protein. *a*, G6 plot for chicken TPI (ref. 9). The distance between two C α carbons at positions *i* and *j* is noted with a black box at point P(*i*, *j*) if it is greater than 2.8 nm, and an open box if less than 0.7 nm (where *i* and *j* represent the amino-acid residue numbers). Pale lines correspond to splice junctions for introns from *Aspergillus* and maize, the dark line represents the *Culex* splice junction site. The hypothetical exons are noted in Roman numerals along the diagonal. An exon encodes a module if the triangle formed by the diagonal and the introns does not contain a dark box²⁴. *b*, Primary structure of a hypothetical ancestral TPI protein (based on ref. 2). Arrows represent β -strands, and boxes α -helices. The splice junction sites for introns from *Culex*, *Aspergillus* and maize are marked by vertical lines. The exons are labelled as in *a*.

this intron in *C. tarsalis*. As shown in Fig. 2, its position is quite precise, separating the long exon III into two modular exons; exon III_A encodes the secondary structure of $\beta\alpha\beta$ (bBc), and exon III_B the highly conserved loop containing part of the active site of the second subunit¹¹. In this regard it should be noted that modular designations are not dictated by secondary structure¹⁰.

The only other insect TPI sequence known is from *Drosophila melanogaster*¹² and it too has a single intron in the coding region, but it separates Gln 179 and Ala 180, a site which is conserved in vertebrates and similar to one in maize (Fig. 2). It is intriguing that the *Culex* intron is at a site that is unique among TPis studied. From the 'introns early' point of view, the absence of the *Culex* intron in another dipteran implies that there were at least three independent events of loss of this intron: one in a plant/fungus ancestor (or one in each of these lines after their divergence), a second in a vertebrate ancestor, and a third in insects after the divergence of *Drosophilidae* and *Culicidae*¹³.

This suggests that intron loss is a common event. Balanced against a drive for stochastic loss there may be pressure on insects with streamlined genomes to retain an intron to house specific enhancers. In contrast to the multiple incidences of loss, the 'introns late' view seemingly requires only a single insertion to explain the presence of the *Culex* intron. Because insertion at a modular boundary is difficult to explain, especially when introns from other TPI genes are considered, the presence of this intron leads us to favour the introns early hypothesis.

Thus, although our data do not firmly resolve the question concerning the origin of introns¹⁴, we do provide evidence supporting the assembly of ancient genes by exon shuffling. The implications for insect phylogeny are interesting, as *Culex* mosquitoes are thought to be primitive dipterans. If the intron observed is indeed ancient, then it should be observed in other mosquitoes, or their ancestors, such as the Hymenoptera (wasps and bees), or Mecoptera (scorpion flies)¹⁵. □

Received 26 August; accepted 12 November 1992.

1. Gö, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1964–1968 (1983).
2. Gilbert, W., Marchionni, M. & McKnight, G. *Cell* **46**, 151–154 (1986).
3. Marchionni, M. & Gilbert, W. *Cell* **46**, 133–141 (1986).
4. Strauss, D. & Gilbert, W. *Molec. Cell. Biol.* **5**, 3497–3506 (1985).
5. Cavellier-Smith, T. *Nature* **315**, 283–284 (1985).
6. Palmer, J. D. & Logsdon, J. M. Jr *Curr. Opin. Genet. Dev.* **1**, 470–477 (1991).
7. Gilbert, W. *Nature* **271**, 501 (1978).
8. Doolittle, W. F. *Nature* **272**, 581–582 (1978).
9. Gilbert, W. *Science* **228**, 823–824 (1985).
10. Gö, M. and Nosaka, M. *Cold Spring Harb. Symp. quant. Biol.* **52**, 915–924 (1987).
11. Banner, D. W. *et al. Nature* **255**, 609–614 (1975).
12. Shaw-Lee, R. L., Lissimore, J. L. & Sullivan, D. T. *Molec. Gen. Genet.* **230**, 225–229 (1991).
13. Futuyama, D. J. *Evolutionary Biology* 2nd edn 323 (Sinauer, Sunderland, MA, 1986).
14. Hickey, D. A., Dunkel, B. F. & Abukashawa, S. M. *J. theor. Biol.* **137**, 41–53 (1989).
15. Daly, H. V., Doyen, J. T. & Ehrlich, P. R. *Introduction to Insect Biology and Diversity* 268 (McGraw-Hill, New York, 1978).

16. Ziegler, R., Whyard, S., Downe, A. E. R., Wyatt, G. R. & Walker, V. K. *Pestic. Biochem. Physiol.* **28**, 279–285 (1987).
17. McNight, G. M., O'Hara, P. J. & Parker, M. L. *Cell* **46**, 143–147 (1986).
18. Old, S. E. & Mohrenweiser, H. W. *Nucleic Acids Res.* **16**, 9055 (1988).
19. Craig, L. C., Pirtle, I. L., Gracy, R. W. & Pirtle, R. M. *Gene* **99**, 217–227 (1991).
20. Brown, J. R., Daar, I. O., Krug, J. R. & Maquat, L. E. *Molec. cell. Biol.* **5**, 1964–1976 (1985).
21. Chia, W., Karp, R., McGill, S. & Ashburner, M. *J. Molec. Biol.* **186**, 689–706 (1985).
22. Glisin, V., Crkvenjakov, R. & Byus, C. *Biochemistry* **13**, 2633–2674 (1974).
23. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
24. Gö, M. *Nature* **291**, 90–93 (1981).

ACKNOWLEDGEMENTS. We thank N. Besansky and A. E. R. Downe for mosquitoes, T. Inalsingh and C. Lyons for DNA and protein sequencing, W. Gilbert for providing materials for figures, and for reviewing the manuscript, and G. R. Wyatt, D. Hickey and J. Oakeshott for helpful suggestions during preparation of the manuscript. Supported by Insect Biotech Canada, a Network of Centres of Excellence funded by the Government of Canada, and the Natural Sciences and Engineering Research Council (Canada).

Inhibition of HIV-1 infectivity by zinc-ejecting aromatic C-nitroso compounds

William G. Rice*, Catherine A. Schaeffer*, Brad Harten*, Francois Villinger†, Terri L. South‡, Michael F. Summers‡, Louis E. Henderson§, Julian W. Bess Jr§, Larry O. Arthur§, J. Steven McDougal||, Sherry L. Orloff||, Jerome Mendeleyev¶ & Ernest Kun¶

* Laboratory of Antiviral Drug Mechanisms, and § AIDS Vaccine Program, Program Resources Inc./DynCorp, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA

† Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia 30322, USA

‡ Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, Baltimore, Maryland 21228, USA

|| Immunology Branch, DHA, NCID, Centers for Disease Control, US Public Health Service, Atlanta, Georgia 30333, USA

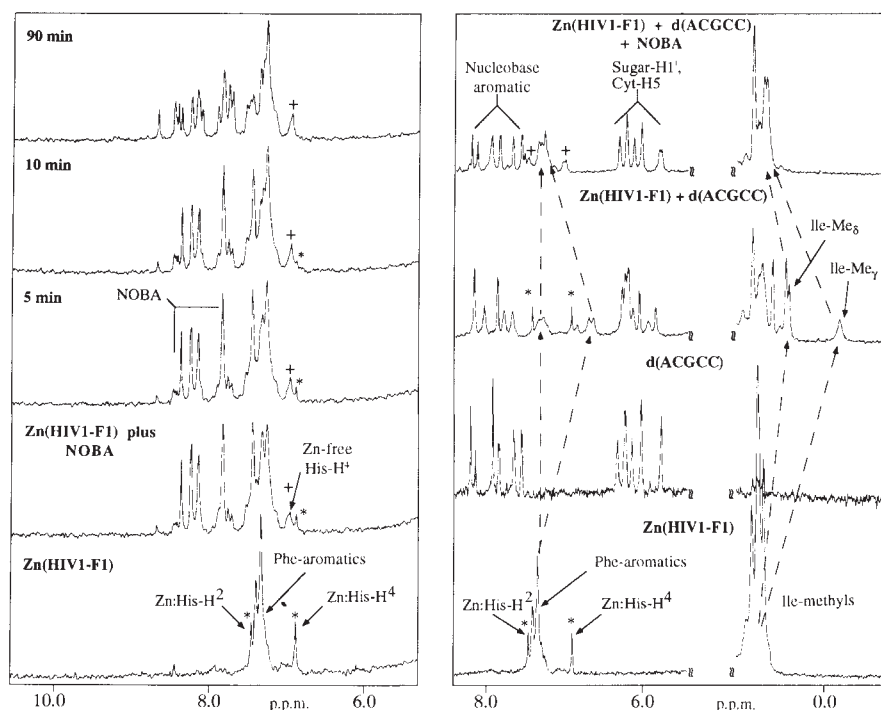
¶ Octamer Inc., and the Laboratory for Environmental Toxicology and Chemistry, The Romberg Tiburon Centers, San Francisco State University, PO Box 855, Tiburon, California 94920, USA

RETROVIRAL nucleocapsid and gag-precursor proteins from all known strains of retroviruses contain one or two copies of an invariant sequence, Cys-X₂-Cys-X₄-His-X₄-Cys^{1,2}, that is populated with zinc in mature particles³. Modification of cysteine or histidine residues results in defective packaging of genomic viral RNA and formation of non-infectious particles⁴⁻⁸, making these structures potentially attractive targets for antiviral therapy^{3,8}. We recently reported that aromatic C-nitroso ligands of poly(ADP-ribose) polymerase preferentially destabilize one of the two (Cys-X₂-Cys-X₂₈-His-X₂-Cys) zinc-fingers with concomitant loss of enzymatic activity^{9,10}, coincidental with selective cytotoxic

action of the C-nitroso substituted ligands on cancer cells¹¹. Based on the occurrence of (3Cys, 1His) zinc-binding sites in both retroviral nucleocapsid and gag proteins and in poly(ADP-ribose) polymerase¹², we reasoned that the C-nitroso compounds may also have antiretroviral effects. We show here that two such compounds, 3-nitrosobenzamide and 6-nitroso-1,2-benzopyrone, inhibit infection of human immunodeficiency virus HIV-1 in human lymphocytes and also eject zinc from isolated HIV-1 nucleocapsid zinc fingers and from intact HIV-1 virions. Thus the design of zinc-ejecting agents that target retroviral zinc fingers represents a new approach to the chemotherapy of AIDS.

Retroviral zinc-fingers with the sequence Cys-X₂-Cys-X₄-His-X₄-Cys (CCHC) bind zinc stoichiometrically and with high affinity^{3-8,13-17} (dissociation constant K_d , about 10^{-12} M; refs 18, 19). Under almost physiological conditions of ionic strength and pH, a 10-fold excess of EDTA removes only 50% of the zinc from the N-terminal CCHC zinc-finger domain of the HIV-1 nucleocapsid protein². To determine whether 3-nitrosobenzamide (NOBA) is capable of ejecting zinc from retroviral-type zinc-fingers, we used NMR to study an 18-residue peptide with sequence corresponding to the N-terminal zinc-finger domain of the HIV-1 nucleocapsid protein, Zn(HIV-1-F1)²⁰. Previously NMR has shown that this peptide binds zinc stoichiometrically and with high affinity¹³, and three-dimensional structural studies indicate that the peptide adopts a structure that is essentially identical to that of the corresponding region in the intact nucleocapsid protein^{3,14,16}. As illustrated by the ¹H-NMR spectrum in Fig. 1 (left panel), addition of NOBA results in the loss of the signals due to zinc-bound histidine (denoted by an asterisk) and the appearance of broad signals representing zinc-free histidine (plus signs)^{13,21}. Other signals in the spectrum are due to NOBA protons. After 90 min, no signals attributable to the zinc-bound histidine could be detected (Fig. 1, upper left), and the signals due to unreacted NOBA were of the same intensity as the reacted NOBA signals, indicating that NOBA reacts stoichiometrically with Zn(HIV-1-F1). Addition of 6-nitroso-1,2-benzopyrone (NOBP) to Zn(HIV-1-F1) gives identical spectral changes, although zinc ejection occurs at a reduced rate (~10-fold).

FIG. 1 NMR of Zn(HIV-1-F1) and NOBA interactions. Left, downfield region of the ¹H NMR spectrum obtained for Zn(HIV-1-F1) (1 mM in D₂O solution, pD=6.2, T=30 °C) (bottom) and upon addition of NOBA (2 mM). The '*' and '+' symbols denote the aromatic proton signals of the zinc-coordinated and zinc-free His9 residue, respectively^{13,14,21}. Upon completion of the reaction, reflected by complete ejection of zinc ($t=90$ min), 50% of the NOBA remained unreacted, indicating a 1:1 reaction stoichiometry. NOBA was synthesized as reported¹⁰. The peptide sequence, VKCFNCGKEGHIARNCR, corresponds to the N-terminal CCHC array of the HIV-LA1 gag polyprotein (amino acids 390-470)²⁰. Right, selected regions of the ¹H NMR spectrum obtained for Zn(HIV-1-F1) (bottom) and a synthetic oligonucleotide with sequence corresponding to a region of the HIV-1 psi-packaging signal, d(ACGCC) (second from bottom). The downfield region of the spectrum shows the signals of the aromatic and ribose H1' protons, and the upfield region contains signals of the methyl group (Me) protons. Dramatic spectral changes occur upon addition of oligonucleotide to Zn(HIV-1-F1), including large upfield shifts of the Phe2 and Ile10 side-chain signals and a downfield shifting and broadening of the guanosine-H8 proton signal (second from top). After incubation with two equivalents of NOBA, the spectral features are characteristic of metal-free peptide and dissociated nucleic acid (top), indicating that NOBA-induced zinc ejection leads to loss of high-affinity nucleic acid-binding function. No direct interaction between NOBA and the



nucleic acid substrate was observed. Similar results were obtained with NOBP.

TABLE 1 Zinc loss resulting from treatment of HIV-1 virions with NOBA

NOBA (μM)*	Incubation time (h)	Zinc (control sample, $\mu\text{g ml}^{-1}$)	Zinc (NOBA-treated sample, $\mu\text{g ml}^{-1}$)	Zinc loss (%)
3,000	2	0.21	0.11	52
6,000	4	0.24	0.04	83

HIV-1 (MN strain), purified and concentrated as described³⁵, was diluted 60 $\times$ in TNE buffer (0.01 M Tris-HCl, 0.1 M NaCl, 1 mM EDTA, pH 7.2) and incubated with NOBA at 37 °C. Virus was then pelleted and washed with TNE buffer to remove weakly bound zinc. Zinc in the viral pellets was determined as described³⁵. No significant loss of viral proteins in the pellet was detected by p24 and gp120 competition radioimmunoassays and Coomassie-stained SDS-polyacrylamide gel electrophoresis.

* Concentrations correspond to molar NOBA:zinc-finger ratios of $\sim 350:1$ (3,000 μM) and 700:1 (6,000 μM).

Retroviral CCHC zinc-fingers participate in several nucleic acid interactions during the course of the virus life cycle, including the specific recognition of the viral genome during budding⁴⁻⁸, packaging of RNA in mature virions^{1,4,7,8,22-26}, and facilitation of reverse transcription^{7,27-31}. Zn(HIV-1-F1) shows sequence-dependent, high-affinity binding to nucleic acids ($K_d = 5 \times 10^{-6}$ M), and the three-dimensional structure of the complex with a 5-residue oligonucleotide, d(ACGCC), has been determined to high atomic resolution²¹. Addition of NOBA (or NOBP) to this CCHC zinc-finger-nucleic acid model complex results in ejection of zinc and dissociation of the complex (Fig. 1, right panel). Identical spectral changes are induced by addition of a 50-fold excess of EDTA²¹. Thus the CCHC arrays of the HIV-1 nucleocapsid and *gag* proteins may be specifically altered so that binding to nucleic acids is prevented.

Both NOBA¹⁰ and NOBP⁹ were tested for their inhibitory effect on HIV-1 infectivity in human lymphocytes. As illustrated in Fig. 2a, the infectivity titre of HIV-1 was reduced by 3-4 orders of magnitude on preincubation with the C-nitroso drugs, whereas there was barely any inhibition of HIV-1 replication when the drugs and virus were diluted separately before addition to the lymphocyte culture. HIV-1 propagation was depressed more effectively by NOBA, consistent with the NMR results, and was also dependent on the length of exposure of the virus to the drug (Fig. 2b). It is important that NOBA is well tolerated by the lymphocytes, with concentrations up to 50 μM (which inactivated 10^5 TCID₅₀ (tissue culture 50% infective dose) of HIV-1) causing no major changes in cell metabolism (Fig. 2c), as determined by a quantitative dye reduction³². Thus, in agreement with our previous results with cancer cells¹¹, only a negligible cytotoxic effect of C-nitroso drugs on normal human cells was detectable. The intracellular lifetime of C-nitroso drugs is short¹¹ (~ 30 min) and, being converted to non-toxic products, a rapid one-hit cellular mechanism may explain low toxicity at concentrations that induce chemotherapeutic responses.

Concentrated suspensions of HIV-1 (60 $\times$, with respect to culture fluid) were also treated with NOBA, which caused a 50-84% loss of viral zinc (Table 1) and a complete loss of infectivity. As zinc-edge X-ray absorption fine-structure spectroscopy has shown that the majority of the zinc in intact retroviruses is coordinated by the CCHC ligands^{3,33}, the ejection of zinc from virions by NOBA is directly attributable to a destabilization of the nucleocapsid CCHC zinc-fingers. The anti-HIV activity observed previously³⁴ for R-NH₂-type ligands of poly(ADP-ribose)polymerase may now be attributed, in part, to destabilization of retroviral CCHC zinc-fingers because R-NH₂ compounds are metabolic precursors of R-NO-type molecules⁹. The correlation of antiretroviral activity of these R-NH₂ drugs in infected cells with their binding to the poly-

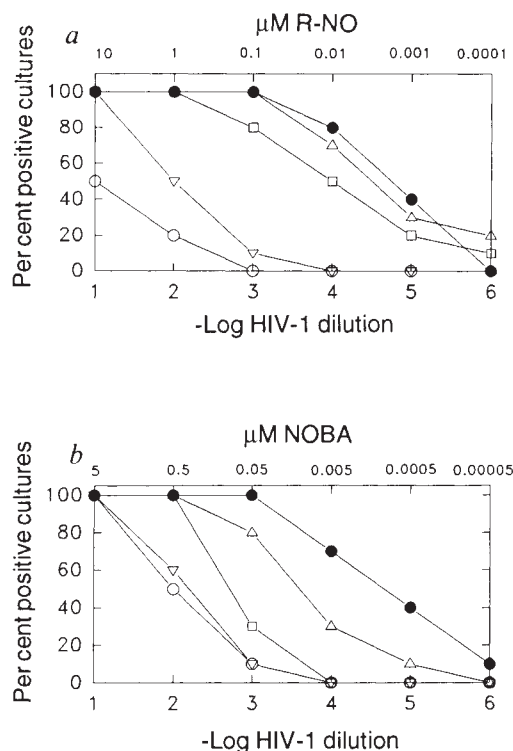


FIG. 2 The effect of C-nitroso drugs on HIV propagation. a, Undiluted stock of HIV-1 (LAV strain; 10^5 TCID₅₀) was incubated in the absence of drugs (●) or with either 100 μM NOBA (○) or 100 μM NOBP (▽) (corresponding

to drug:zinc-finger molar ratios of $\sim 1,400:1$) for 30 min at 22 °C. Serial 10-fold dilutions were made, and the dilutions were added to phytohaemagglutinin-stimulated peripheral blood lymphocytes (PHA-PBL) in microcultures (1×10^5 cells per culture, 10 cultures per dilution). Viral replication was monitored by detection of viral antigens in the culture supernatants after 9 d of culture³⁶, and cultures were scored positive if their absorbance at 490 nm in the antigen-capture ELISA was >3 s.d. above the mean absorbance of 10 uninfected cultures. Alternatively, the drugs (NOBA, Δ, or NOBP, □) and HIV-1 were kept in separate tubes, diluted separately, and the dilutions combined just before addition to cell cultures. The virus (-log HIV-1 dilution) and drug (μM R-NO) concentration in cultures from 'preincubated' and 'non-premixed' samples were identical as indicated at the top abscissa; therefore the reduction in virus titre is the result of a direct action of C-nitroso drugs on the HIV-1 virus. b, HIV-1 stock was preincubated with NOBA (50 μM , corresponding to a NOBA:zinc-finger molar ratio $\sim 700:1$) at 37 °C for 5 min (Δ), 15 min (□), 30 min (▽), or 60 min (○). Mixtures were serial 10-fold diluted and infectivity monitored as described for a. In the controls (●) HIV-1 virus was titrated in the absence of drug. c, Uninfected PHA-PBL (10^6 per ml) were treated with increasing doses of NOBA for 24 h and the metabolic activity of the cells was measured by the microculture tetrazolium assay³². The rate of reduced dye formation in the absence of NOBA was taken as 100% and experimental values were normalized to this control value.

merase may indicate that they work by mechanisms like those operating in cells containing amplified oncogenes¹¹.

Our results anticipate the application of R-NO-type C-nitroso drugs for AIDS chemotherapy, identifying the essential nucleocapsid zinc-fingers of the virus as new targets. As the rate of zinc ejection and antiviral activity of R-NO drugs depends on the molecular structure of R, variants can be synthesized to optimize efficacy by altering the R group. □

Received 23 July; accepted 26 November 1992.

- Henderson, L. E., Copeland, T. D., Sowder, R. C., Smythers, G. W. & Oroszlan, S. *J. Biol. Chem.* **256**, 8400–8406 (1981).
- Summers, M. F. *J. Cell Biochem.* **45**, 41–48 (1991).
- Summers, M. F. *et al. Protein Science* **1**, 563–574 (1992).
- Aldovini, A. & Young, R. A. *J. Virol.* **64**, 1920–1926 (1990).
- Lever, A., Gottlinger, H., Haseltine, W. & Sodroski, J. *J. Virol.* **63**, 4085–4087 (1989).
- Gorelick, R. J., Henderson, L. E., Hanser, J. P. & Rein, A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8420–8424 (1988).
- Meric, C. & Goff, S. P. *J. Virol.* **63**, 1558–1568 (1989).
- Gorelick, R. J. *et al. Virol.* **64**, 3207–3211 (1990).
- Buki, K. G., Bauer, P. I., Mendeleyev, J., Hakam, A. & Kun, E. *FEBS Lett.* **290**, 181–185 (1991).
- Kun, E., Mendeleyev, J. & Rice, W. G. US Patent ser. no. 07/780,809, pending.
- Rice, W. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 7703–7707 (1992).
- Gradwohl, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 2990–2994 (1990).
- South, T. L., Kim, B. & Summers, M. F. *J. Am. chem. Soc.* **111**, 295–396 (1989).

- Summers, M. F., South, T. L., Kim, B. & Hare, D. R. *Biochemistry* **29**, 329–340 (1990).
- South, T. L., Blake, P. R., Hare, D. R. & Summers, M. F. *Biochemistry* **30**, 6342–6349 (1991).
- South, T. L. *et al. Biochemistry* **29**, 7786–7789 (1990).
- Fitzgerald, D. W. & Coleman, J. E. *Biochemistry* **30**, 5195–5201 (1991).
- Green, L. M. & Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4047–4051 (1989).
- Green, L. M. & Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6403–6407 (1990).
- Myers, G., Korber, B., Berzofsky, J. A. & Smith, R. F. *Human Retroviruses and AIDS* (Los Alamos National Laboratories, New Mexico, 1991).
- South, T. L. & Summers, M. F. *Protein Science* **2**, 3–19 (1993).
- Smith, B. J. & Baily, J. M. *Nucleic Acids Res.* **7**, 2055–2072 (1979).
- Chen, M.-J., Garon, C. F. & Papas, T. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4961–4967 (1980).
- Karpel, R. L., Henderson, L. E. & Oroszlan, S. *J. Biol. Chem.* **262**, 4961–4967 (1987).
- Meric, C., Darlix, J.-L. & Spahr, P.-F. *J. molec. Biol.* **173**, 531–538 (1984).
- Dupraz, P., Oertle, S., Meric, C., Damay, P. & Spahr, P.-F. *J. Virol.* **64**, 4978–4987 (1990).
- Bieth, E., Gabus, C. & Darlix, J.-L. *Nucleic Acids Res.* **18**, 119–127 (1990).
- Prats, A.-C. *et al. EMBO J.* **7**, 1777–1783 (1988).
- Prats, A.-C. *et al. Nucleic Acids Res.* **19**, 3533–3541 (1991).
- Cornille, F. *et al. Int. J. Peptide Protein Res.* **36**, 551–558 (1990).
- Meric, C. & Spahr, P.-F. *J. Virol.* **60**, 450–459 (1986).
- Mosmann, T. *J. Immun. Meth.* **65**, 55–63 (1983).
- Chance, M. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 10041–10045 (1992).
- Cole, G. A. *et al. Biochem. biophys. Res. Commun.* **180**, 504–514 (1991).
- Bess, J. W. *et al. J. Virol.* **66**, 840–847 (1992).
- McDougal, J. S. *et al. J. Immun. Meth.* **76**, 171–183 (1985).

ACKNOWLEDGEMENTS. This work was supported by the Octamer Research Foundation, the NIH, and the National Cancer Institute, DHHS, under contract with PRI/DynCorp. The contents of the publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Structure of a molecular chaperone from a thermophilic archaebacterium

Barry M. Phipps^{*,†}, Dieter Typke^{*}, Reiner Hegerl^{*}, Susanne Volker^{*}, Angelika Hoffmann[‡], Karl O. Stetter[‡] & Wolfgang Baumeister^{*}

^{*}Max-Planck Institute for Biochemistry, W-8033 Martinsried, Germany

[‡]Department of Microbiology, University of Regensburg, W-8400 Regensburg, Germany

WE recently reported that a wide range of thermophilic archaebacteria have a novel ATPase complex which accumulates to high levels upon heat shock and may be a new type of molecular chaperone related to the chaperonins¹. Striking similarities between the complex, referred to here as the 'thermosome', and the subsequently characterized chaperones TF55 of the thermophilic archaebacterium *Sulfolobus shibatae*² and TCP-1 from eukaryotic cytosol^{2–5} suggest that these proteins belong to a single family. Here we determine the three-dimensional structure of the thermosome from *Pyrodicticum occultum* by random conical tilt reconstruction from electron micrographs. The reconstruction reveals a complex consisting of two rings of eight subunits each, stacked face-to-face. The subunits are kidney-shaped and composed of at least two domains. In the centre of the thermosome is a large cavity of about 6.7 nm diameter; the opening of this cavity on each face of the complex is partially blocked by a mass which appears to be weakly connected to the eight-membered ring.

When negatively stained and examined by transmission electron microscopy, the thermosome adopts two distinct orientations with respect to the carbon film to give characteristic end and side views¹ (Fig. 1a, c, d). End views comprise a ring with eightfold radial (pseudo)symmetry enclosing a central stain-filled region. A few specimens (<10%) appear to have ninefold symmetry. Side views are barrel-shaped with four parallel striations, two thick ones in the middle and two thin ones at the ends, and appear to be twofold-symmetric both up-and-down and side-to-side. The thermosome from the hyperthermophile *Pyrodicticum occultum* is composed of equal amounts of two

polypeptides of M_r 56,000 and 59,000 (56K and 59K) and is an extremely thermophilic and thermostable ATPase¹. It is one of the principal cytoplasmic proteins and its relative abundance increases dramatically on heat shock (shift from 102 °C to 108 °C), which correlates with the acquisition of thermotolerance (A.H. and K.O.S., unpublished data). A protein related to the *Pyrodicticum* thermosome is found in a wide range of thermophilic archaebacteria¹.

We have proposed¹ that the thermosome is a new type of molecular chaperone^{6–8} related to the chaperonin-60 proteins^{6,9,10} (cpn60s), which have a characteristic structure consisting of two stacked seven-membered rings of 60K subunits and are superficially similar in the electron microscope⁹. Subsequently, proteins have been identified which possess chaperone activity and are very similar in structure to the thermosome^{2–5,11}, giving nearly identical eight- and ninefold symmetric ring-shaped end views and striated rectangular side views in the electron microscope. TF55 is a large complex composed of a single 55K polypeptide. It has thermophilic ATPase activity and is abundant at physiological growth temperatures; it accumulates to higher levels when the temperature is upshifted, concomitant with acquisition of thermotolerance^{2,12}. The amino-terminal sequences of five tryptic peptides derived from the thermosome 56K subunit shared significant homology with regions of TF55 and of these, four were more than 70% identical (data not shown). The TCP-1 complex is an ATPase and exists as a heteromeric assembly (M_r 800–950K) of several polypeptides in the 53–72K range, including the conserved 58K t-complex polypeptide-1 (refs 3–5). It has been suggested³ that the side views actually represent a dramatically altered conformation of the TCP-1 complex, but this is almost certainly a shift in preferred orientation from end-on to side-on, as seen for other macromolecular complexes¹³. TCP-1 is significantly homologous with TF55 (ref. 2), and slightly similar to the cpn60s^{4,14,15}. Thus it seems that the TCP-1 complex in eukaryotic cytoplasm, with the thermosome and TF55 in thermophilic archaebacteria, constitute a family of closely related chaperones having features in common with the chaperonins.

The three-dimensional structure of the *Pyrodicticum* thermosome was determined by random conical tilt reconstruction¹⁶ from electron micrographs of negatively stained complexes in end-on orientation (Fig. 1a). Micrographs of specimens tilted to 60° were recorded using dynamic focusing¹⁷ to correct for the steep focus gradient encountered at high tilt angles (Fig. 1b). Two independent reconstructions gave essentially identical results and only one is documented here. Horizontal and vertical

[†]Present address: Institute for Biological Sciences, National Research Council, Ottawa, Ontario, Canada K1A 0R6.

sections through the reconstruction are shown in Fig. 2. The bottom half of the reconstruction (closest to the carbon film) has the same features as the top, but shows less contrast and fades out more quickly, presumably as a result of deeper embedding in stain or flattening upon adsorption to the carbon film (Fig. 2a). In addition, the molecule has a distinct handedness and that of the bottom half is opposite that of the top half. The twofold symmetry of side views implies that the top and bottom of the complex are equivalent, so the molecule must consist of two equal halves interacting at the same face. We therefore replaced the bottom half of the reconstruction with an inverted copy of the top half (Fig. 2b), which was probably better preserved as it had not been subjected to adsorptive forces from contact with the carbon film. Vertical sections through this reconstruction are in accord with the side views (Fig. 2c). The final three-dimensional model is shown in Fig. 3 as a surface-shaded solid and as a see-through network.

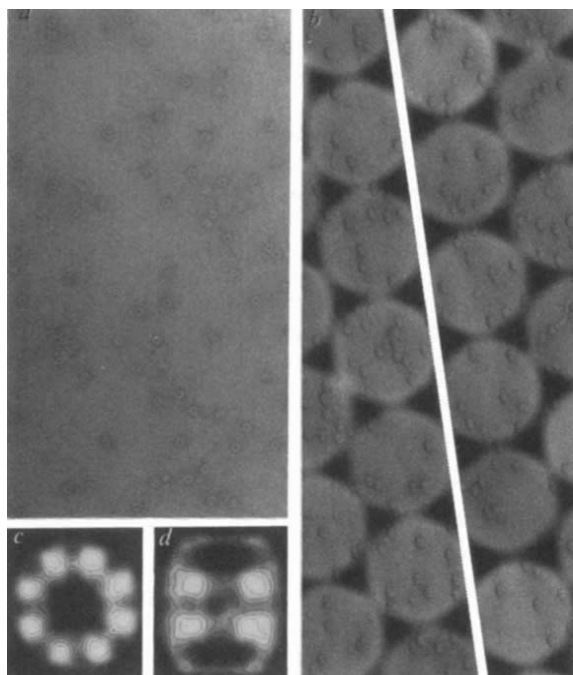


FIG. 1 Transmission electron microscopy and image processing of negatively-stained *Pyrodicticum* thermosomes. *a, b*, Extracts of micrographs used for three-dimensional reconstruction. *a*, Flood-beam image of untilted specimen. Most molecules are in the end-on orientation (rings), but some side views (striated 'barrels') are also seen. *b*, Spot-step image of the same field recorded with the specimen at a tilt angle of 60°, using dynamic focusing to achieve a narrow focus range over the entire micrograph. The white line indicates the tilt axis. Magnification, 103,000 $\times$. *c, d*, Correlation averages of the *Pyrodicticum* thermosome. End and side views were separately aligned and averaged using correlation methods. *c*, Average of 264 end views. Diameter of ring and central stain-filled hole are 16 nm and 5 nm respectively. *d*, Average of 284 side views. Maximum height and width are 16.5 nm and 15.5 nm respectively. The two middle striations (each consists of right and left stain-excluding peaks) have a combined height of 10.5 nm. Panels *c* and *d* are reproduced from ref. 1 with permission.

METHODS. *Pyrodicticum occultum* thermosomes were purified as described¹, except that cells were grown to late exponential phase at 102 °C and then shifted to 108 °C for 3 h, resulting in lysates highly enriched in the protein. The particles were stained on grids with 2% uranyl acetate, and examined in a Philips CM12 electron microscope operated in low-dose mode at 100 kV. Pairs of micrographs were recorded at nominal tilt angles of 60° and 0° and a magnification of 33,600 $\times$. The micrograph of the tilted specimen was taken using spot-step illumination combined with dynamic focusing, carried out via external TEMDIPS computer control of the microscope. Total electron dose was ~ 300 electrons nm⁻². Fast scanning direction was parallel to the tilt axis and the focus was incremented by 290 nm between rows of spots. Mean defocus was about -400 nm. A micrograph of the untilted specimen was subsequently recorded using flood-beam illumination.

The model comprises two identical rings of eight kidney-shaped structures, each of which is separated from its neighbour (Figs 2, 3) and probably corresponds to a single 56K or 59K subunit, given that the complex is likely to be a hexadecamer of 56K and 59K monomers in a 1:1 ratio¹. The subunits are asymmetric and oriented identically in the ring, with each ring

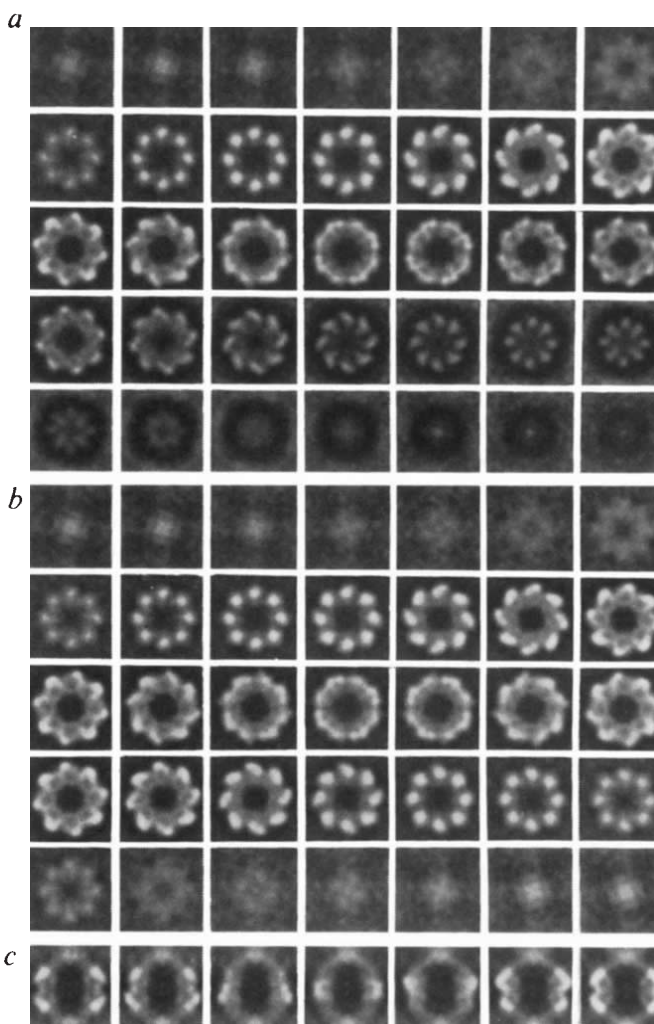


FIG. 2 Sections through the three-dimensional reconstruction of the thermosome. *a, b*, Horizontal sections. *a*, Fourfold rotational symmetrization was applied around an axis perpendicular to the sections (*z* axis). The presence of two dissimilar subunits in the thermosome implies that the true rotational symmetry is probably fourfold. Distance between successive sections is 0.45 nm. *b*, As *a*, except that the bottom half of the reconstruction was discarded and replaced with the top half by substituting each section with the mirror image of the corresponding section from the top half. *c*, Radial vertical sections through completed reconstruction. Sections are parallel to the *Z* axis and pass through the centre of the model. The first section is parallel to the *x* axis and the angular increment between successive sections is 7.5°. Dimensions of all sections are 19.6 nm $\times$ 19.6 nm.

METHODS. Reconstruction was carried out using the random conical tilt method¹⁶, implemented in the EM software package²⁹. Corresponding micrographs of tilted and untilted thermosomes were digitized and filtered. Individual thermosome end views (238) in the 0° image were selected interactively on a graphics display and aligned via cross-correlation techniques. A linear transformation relating the positions of tilted and untilted complexes in the two images was calculated and used to extract a set of projections from the tilt image. Azimuthal angles of the projection directions were determined from in-plane rotation angles found during alignment and a reconstruction computed by weighted back-projection, using an exact filter function based on the sinc-function. Each input projection image was aligned to a corresponding projection through the reconstruction. The new alignment was accepted as long as the corresponding particle in the 0° image was still well aligned. The new set of projections was used to recompute the reconstruction, significantly improving the final model.

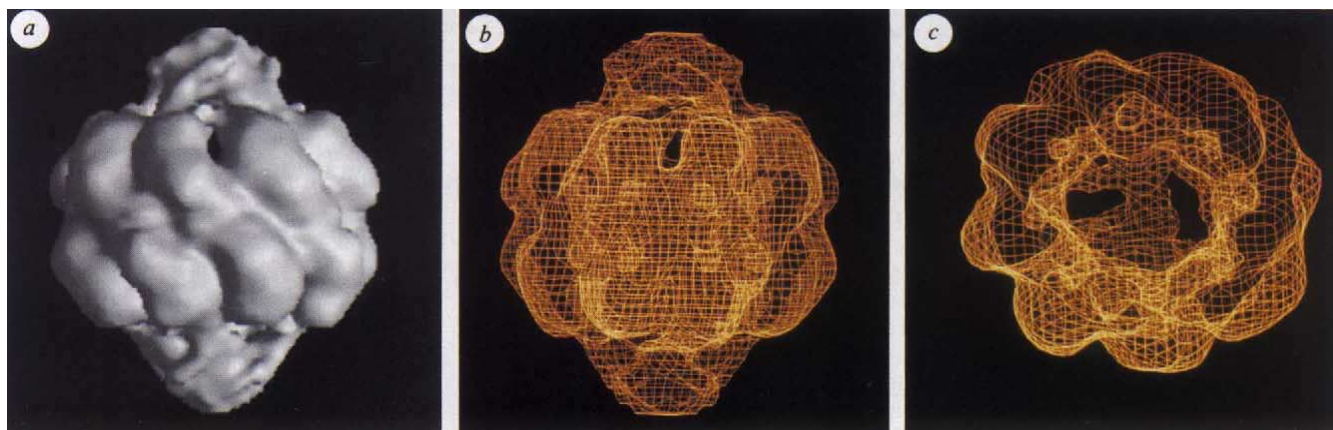


FIG. 3 Three-dimensional (3-D) models of the thermosome. *a*, Surface-shaded solid model, viewed from the side of the molecule. Two rings containing eight kidney-shaped subunits each are stacked such that equivalent faces are in contact. The rings enclose a central cavity which is partially capped at both ends by a mass with narrow connections to the rings. Height of the model is 19 nm, diameter is 16.5 nm, and combined height of the two stacked rings is 10.5 nm. *b*, *c*, Network representations of the 3-D model: *b*, side-on view; *c*, the top part of the molecule has been cut away and the view is into

having two different faces; equivalent faces are in contact and each subunit pairs up with a subunit in the opposing ring. The putative monomer can be divided into two or three major domains. The striations seen in electron microscopic side views of cpn60s are parallel to the plane of the seven-membered rings; two striations correspond to each ring, suggesting that the 60K subunits consist of at least two domains^{18,19}. Different domains may have separate functions, for example, ATP hydrolysis or peptide binding^{20,21}.

A conspicuous feature of the reconstruction is a large central cavity with corrugated walls and a mean diameter of 6.7 nm. A cavity is also seen in a less detailed reconstruction of the bacterial cpn60 groEL²². A central cavity is likely to be important functionally. Chaperones are believed to bind noncovalently to transiently exposed surfaces presented by unfolded or partially folded proteins^{8,23,24}. Controlled release, involving ATP hydrolysis, leads to correct folding or assembly. There is evidence²⁵ that the binding site(s) for unfolded polypeptides lies in the cavity of groEL²⁶. The dimensions of the thermosome cavity would accommodate a large polypeptide chain, either in the folded or unfolded state.

The opening of the cavity on each face of the complex is partially occluded by a solid mass which appears to be joined to the ring by narrow extensions. The mass is unlikely to be an artefact of the reconstruction because a steep density gradient suggests the presence of a protein-stain boundary and the appearance of side views (Fig. 1*d*) agrees with our interpretation. The two eight-membered rings probably correspond to the two thick striations in the middle of the side view, and the

the large central cavity; mean diameter of the cavity is 6.7 nm.

METHODS. The 3-D model was generated by thresholding sections through the reconstruction using the criterion that the volume of the final model should correspond to 100% of the expected mass of the thermosome, assuming a density of 1.3 gm cm^{-3} . Assuming that it is composed of eight 56K and eight 59K subunits, the total mass of the thermosome would be 920K (ref. 1). Because of uncertainty in the size and shape of the masses 'capping' the central cavity, the height of the model is only approximate.

masses over the cavity openings to the thin striations at the top and bottom. The height and diameter of the pair of stacked rings ($10.5 \text{ nm} \times 16.5 \text{ nm}$) match the dimensions of the two middle striations (Fig. 1*d*). Thus, as in cpn60s, the striations are parallel to the rings. The actual shape of the mass is uncertain because of the 'missing cone' effect, which results from a lack of data at tilt angles $>60^\circ$, and its lower density in the reconstruction. This, and differences in stain distribution between end and side views, may explain why the mass appears compact in the reconstruction whereas side views show a narrow stripe. The connectivity of the mass to the ring is questionable as the width of the connecting regions is below the resolution of the reconstruction.

The mass may be formed by coalescence of domains extending up and towards the centre of the ring from every subunit in the ring, or from every second subunit (for example, 59K subunits). Alternatively, it might be equivalent to chaperonin-10 (cpn10), which binds as a seven-membered ring over one mouth of the cavity of the cpn60 tetradecamer^{25,27}; the bacterial cpn10 groES is required for ATP hydrolysis-dependent folding of polypeptides on the (presumably cavity-exposed) surface of groEL^{8,24,26,28}. However, silver staining of purified thermosomes fails to detect any additional polypeptides. A third possibility is that the cpn10 and cpn60 equivalents are integrated into a single polypeptide, with the cpn10-related domains poised over the mouth of the cavity. To our knowledge, this is the first detailed structure of a chaperonin or chaperonin-like protein, which should provide a structural basis for further investigation into the mechanism of chaperone function. □

Received 6 August; accepted 30 November 1992.

1. Phipps, B. M., Hoffmann, A., Stetter, K. O. & Baumeister, W. *EMBO J.* **10**, 1711-1722 (1991).
2. Trent, J. D., Nimmeggern, E., Wall, J. S., Hartl, F.-U. & Horwich, A. L. *Nature* **354**, 490-493 (1991).
3. Gao, Y., Thomas, J. O., Chow, R. L., Lee, G.-H. & Cowan, N. J. *Cell* **69**, 1043-1050 (1992).
4. Lewis, V. A., Hynes, G. M., Zheng, D., Saibil, H. & Willison, K. *Nature* **358**, 249-252 (1992).
5. Yaffe, M. B. *et al. Nature* **358**, 245-248 (1992).
6. Gething, M.-J. & Sambrook, J. *Nature* **355**, 33-45 (1992).
7. Ellis, R. J. *Nature* **328**, 378-379 (1987).
8. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. *Nature* **342**, 884-889 (1989).
9. Hemmingsen, S. M. *et al. Nature* **333**, 330-334 (1988).
10. Ellis, R. J. & van der Vies, S. M. *A. Rev. Biochem.* **60**, 321-347 (1991).
11. Ellis, R. J. *Nature* **358**, 191 (1992).
12. Trent, J. D., Osipiuk, J. & Pinkau, T. *J. Bact.* **172**, 1478-1484 (1990).
13. Baumeister, W. *et al. FEBS Lett.* **241**, 239-245 (1988).
14. Ellis, R. J. *Science* **250**, 954-959 (1990).
15. Gupta, R. S. *Biochem. Int.* **4**, 833-841 (1990).
16. Rademacher, M. *J. Electron Microsc. Techn.* **9**, 359-394 (1988).

17. Zemlin, F. *J. Electron Microsc. Techn.* **11**, 251-257 (1989).
18. Hutchinson, E. G., Tichelaar, W., Hofhaus, G., Weiss, H. & Leonard, K. R. *EMBO J.* **8**, 1485-1490 (1989).
19. Viitanen, P. V. *et al. J. Biol. Chem.* **267**, 695-698 (1992).
20. Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. *Nature* **346**, 623-628 (1990).
21. Rippmann, F., Taylor, W. R., Rothbard, J. B. & Green, N. M. *EMBO J.* **10**, 1053-1059 (1991).
22. Carrascosa, J. L., Abella, G., Marco, S. & Carazo, J. M. *J. struct. Biol.* **104**, 2-8 (1990).
23. Flynn, G. C., Pohl, J., Flocio, M. T. & Rothman, J. E. *Nature* **353**, 726-730 (1991).
24. Martin, J. *et al. Nature* **352**, 36-42 (1991).
25. Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F.-U. *EMBO J.* **11**, 4757-4765 (1992).
26. Creighton, T. E. *Nature* **352**, 17-18 (1991).
27. Saibil, H., Dong, Z., Wood, S. & auf der Mauer, A. *Nature* **353**, 25-26 (1991).
28. Viitanen, P. V. *et al. Biochemistry* **29**, 5665-5671 (1990).
29. Hegerl, R. & Altbauer, A. *Ultramicroscopy* **9**, 109-116 (1982).

ACKNOWLEDGEMENTS. We thank H. Deutzmann for peptide sequencing and J. Peters for the sequence comparisons. B.M.P. was the recipient of a Medical Research Council of Canada postdoctoral fellowship.

Of mice and men (and cows and cats)

A compendium of the chromosomal locations of more than 300 human genes and their homologues in three mammalian species will be a boon in comparative mapping studies.

LAST year was the year of the micro-satellite, most notably manifested in the mapping of the human^{1,2} and mouse³ genomes using an abundant supply of these highly polymorphic markers. Microsatellites represent alterations in the length of short nucleotide repeats throughout the genome, and have many advantages over the more conventional but relatively uninformative restriction fragment length polymorphisms (RFLPs). They are amenable to a simple and easily automated analysis, and seem to be evenly spread throughout the genome, which gives them a distinct edge over other classes of repetitive sequence markers. The rapidly expanding supply of these 'second generation' markers has already led to a stunning increase in the density of landmarks strung throughout the human and mouse genomes — for instance, more than 1,000 mapped microsatellites were reported in the human genome^{1,2} towards the end of 1992. Before too long, investigators will probably have at their disposal maps with evenly spaced landmarks whose average spacing is 1 centimorgan (about 1 megabase) or less⁴.

But these new markers have one big drawback: in general they are anonymous stretches of DNA and do not represent coding sequences of genes, which limits their usefulness in drawing comparisons between the genomic organization of different organisms. Coding

sequences usually exhibit less sequence divergence and are therefore not as likely to manifest polymorphisms as non-coding stretches. Ideally, in the next few years the dense microsatellite maps will be integrated with more established compilations of known genes that have been mapped with precision not only in humans, but increasingly in other species as well.

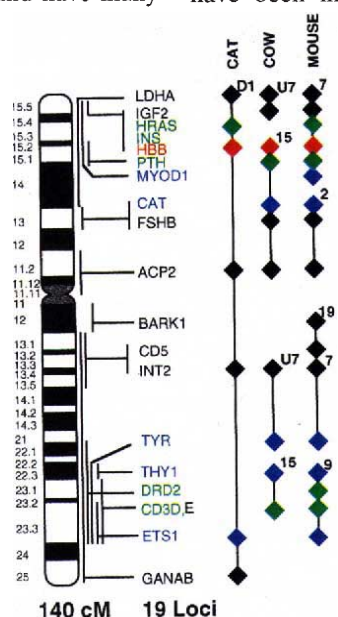
Writing in this month's *Nature Genetics*, Stephen O'Brien and his colleagues announce a major step in this direction⁵. They have compiled a total of 321 cloned and mapped human genes that will be a valuable series of anchor loci for comparative mapping purposes. O'Brien *et al.* have chosen three other species — mouse, cow and cat — to illustrate the utility of these genes. Inclusion of the first animal was mandatory, while the other two are becoming increasingly popular in genetic mapping studies (although to date the number of loci mapped in them is less than a

karyotype analysis. Clusters of gene loci that are physically linked (or syntenic) on a chromosome are often, but not always, conserved through evolution. Identification of such exceptions to the rule plays a large part in determining the patterns of evolutionary divergence. For example, although genes mapped to human chromosome 12 are conserved on single chromosomes in cats and cattle, they are dispersed on at least four different chromosomes in mice⁷. Nonetheless, the mapping of murine genes usually allows accurate predictions of where their human homologues might be, with important consequences for the identification of animal models of human diseases.

To complement the new maps using microsatellites (also termed type II markers^{5,7}), O'Brien *et al.* have selected 321 known genes (type I markers) which will facilitate comparisons between species and provide a framework to monitor the progress of mapping projects in various organisms relying on the more informative microsatellite markers. In time, the two will converge to form the ultimate genetic map. The 321 genes chosen to compile the map have all been cloned and, where possible, are evenly distributed along the chromosomes, giving an average spacing of 5 and 10 centimorgans in the mouse and human genomes, respectively. After choosing as many established human and mouse 'reference loci' (as suggested by gene mapping committees), the authors selected genes in areas subject to recombination between species as well as other loci of particular medical significance. Although the final compilation suffers from some deficiencies (such as uneven spacing) and does not place a premium on practical considerations (including sequence availability and amplification criteria), it is a suitable starting point to accelerate the pace of gene mapping in some mammals, and perhaps to launch projects in others that have been neglected so far.

Kevin Davies

Kevin Davies is Editor of Nature Genetics.



Representative gene map of the human chromosome 11 showing the syntenic relationships of these loci in mice, cats and cattle. Colours show reference loci⁵.

tenth that in humans or mice). Together, these four species represent four different mammalian orders, primates, rodents, artiodactyls and carnivores.

There are essentially two reasons for producing such comparative maps. First, the defined locations of these genes of known function are crucial not only for human geneticists (for elucidating the potential connection with hereditary disease loci, for example) but also for such people as veterinarians and breeders, in that the identification of mapped genes and traits may have important economic or agricultural ends. That point has been underscored lately with the elucidation of the molecular basis of severe genetic disorders in horses, cattle and pigs⁶.

The second reason for comparing different mammalian gene maps is to shed light on the genomic organization and evolution of related species with a resolution that cannot be achieved by

Also in this month's *Nature Genetics*: disruption of the PAX-3 paired box gene by a chromosomal translocation in the solid tumour, alveolar rhabdomyosarcoma; defects in the neurofibromin gene associated with melanoma and neurofibrosarcoma; and a surprising variety of phenotypes resulting from flaws in the retinal protein, peripherin.

1. NIH/CEPH Collaborative Mapping Group *Science* **258**, 67–86 (1992).
2. Weissenbach, J. *et al.* *Nature* **359**, 794–801 (1992).
3. Dietrich, W. *et al.* *Genetics* **131**, 423–447 (1992).
4. Goodfellow, P. N. *Nature* **359**, 777–778 (1992).
5. O'Brien, S. *et al.* *Nature Genet.* **3**, 103–112 (1993).
6. Womack, J. *Nature* **360**, 108–109 (1992).
7. O'Brien, S. *J. Curr. Op. Genet. Dev.* **1**, 105–111 (1992).